Individual and Combined AI models for Complicated Predictive Forest Type Mapping Using Multisource GIS Data

By

Zhi Huang

A thesis submitted for the degree of Doctor of Philosophy of

The Australian National University

School of Resources, Environment and Society

The Australian National University

April, 2004

DECLARATION

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or the other institute of higher learning, except where due acknowledge is made in the text.

Signed

黄智

Zhi Huang

Date: 11/05/04

Acknowledgments

I wish to express my most sincere thanks to the members of my supervisor committee, Dr Brian, G. Lees (ANU), Dr Brian, J. Turner (ANU), and Dr Shawn, W. Laffan (now at UNSW), for their continuous support and invaluable help to my research. Without their tireless efforts and encouragement this thesis is impossible. Their enthusiastic help during the past three years at all aspects of my life are the most important to the successful accomplish of this thesis.

I also wish to thank the colleagues and students at the school of Resources, Environment and Society, ANU. They helped create a friendly and supportive environment for my research and work. I really enjoy the jokes and discussion at teatime every day.

Dr Jeff Wood from the Statistical Consulting Unit of the Graduate School, ANU gave me a lot of consultations. I want to thank him for his help.

Many people have offered me help during my study at ANU, which regrettably cannot been listed here. But their support is appreciated and remembered.

Finally, I want to personally thank my beloved wife Mollie Zeng. We have been through many difficult times, but she has never stopped encouraging and supporting me. She deserves the success of this thesis.

ABSTRACT

This thesis describes a study of forest type mapping using multisource GIS data. The study first investigated the effectiveness of three individual Artificial Intelligence (AI) models including a Decision Tree, an Artificial Neural Network, and a model based on Dempster-Shafer's theory in mapping complicated forest types. The study then developed a new strategy to combine these AI models and examined the advantages of the combination strategy in forest type mapping. Meanwhile, data errors were identified in the study and the modes and effectiveness of the individual and combined AI models in handling the data errors were evaluated. The study also developed Fuzzy Expert Systems for forest type mapping and examined the advantages.

The study found that the three individual AI models were fairly good classifiers for complicated forest type mapping, among them the Decision Tree achieved the best overall and Kappa accuracies. However, it is shown that none of them achieved the best user's accuracies and producer's accuracies on all of the forest types, and they had quite different characteristics in predicting spatial patterns and handling data errors. This is believed to be because of the different principles the three AI models are based on.

On the other hand, the study indicated that the combination strategy was effective and efficient in mapping complicated forest types. The strategy was able to not only improve classification performance and handle data errors effectively but also provide an estimation of prediction confidence that is impossible by using individual classifiers. Several methods including the majority voting system, Dempster-Shafer's theory (Dempster's rule of combination), simple statistical functions, and fuzzy set theory were used for the combination strategy. Two combination stages were subsequently implemented. Among the combined AI models, vote7 at the second combination stages achieved the best overall classification performance, with an increase of over 7% in overall accuracy and an increase of 9% in Kappa accuracy for the forest types from the Decision Tree. The subsequent Z-test shows a significant difference between vote7 and the Decision Tree at the 90% confidence level.

In addition, this study showed that building Fuzzy Expert Systems directly from learning samples was cost-effective and avoided the knowledge acquisition "bottle neck" problem. Applying the Fuzzy Expert Systems to forest type mapping

demonstrated that they were capable classifiers and had the advantages of enhancing comprehensibility, handling classification uncertainty, and providing explanations behind the classification process. However, they also have the limitations of consuming much time and demanding many resources.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

Chapter 1

INTRODUCTION

Predictive modelling for forest mapping is an important innovation for forest management. However, the overall predictive accuracy of forest mapping is disappointing, especially where forest types are discriminated at the Anderson *et al.* (1976) level III (Skidmore & Turner, 1988). This is due to the many uncertainties and great difficulties involved in the mapping process. The goal of this study is to improve the predictive accuracy of forest mapping by using models that can effectively handle these uncertainties and difficulties.

According to Gale (1972), there are three sources of inexactness in classification procedures. First, there is inexactness due to insufficient information, which is associated with the quantity and quality of data. Second, there is inexactness due to the neutrality of predicates, which is associated with the complexity of natural phenomena. Third, there is inexactness due to the effects of secondary qualities, which is associated with observer subjectivity. There is also another source of inexactness not mentioned by Gale (1972), which is associated with the imperfectness of classifiers or predictive models. Together, these four sources of uncertainty pose great difficulty in the mapping of complicated forest ecosystems. However, they have also inspired an enormous body of work, including this thesis, to develop new strategies for effectively handling these uncertainties and improving classification performance.

The following sections describe forest mapping and its uncertainties and difficulties in detail, review past experiences of forest mapping, introduce a new strategy developed in this study, and present the potential advantages of this strategy. These are followed by the objectives and outline of this study.

1.1 FOREST MAPPING AND ITS UNCERTAINTIES AND DIFFICULTIES

There are four sources of uncertainty in the application of forest mapping. The first is associated with the complexity of natural forest ecosystems. A natural forest ecosystem is a continuum in which the boundary between forest types is vague rather than crisp.

The process of forest mapping is to separate the continuum into discrete forest classes which, as pointed out by Lees (1996a), leads inevitably to the generation of errors of omission and commission and results in an upper limit on predictive accuracy. But, when classification is the only appropriate alternative, an effective and suitable classification system needs to be chosen for forest mapping. This leads to the second source of the uncertainties in forest mapping, which is associated with observer subjectivity.

Currently, there are many vegetation (forest) classification systems in Australia and worldwide (e.g., Gillison & Anderson, 1981; Sun *et al.*, 1996). All of them are subjective to some extent and, therefore, each of them is suitable only for certain applications. The definitions of class boundaries are also highly subjective, especially when trying to separate easily confused forest types. For example, the decision of how much variation within class was acceptable and at what point did a variation become a separate class, as well as the different methods of classification, could lead to many different classifications of the same area (McIntosh, 1978). Another factor that may affect the magnitude of the uncertainty is the mapping unit. Forest ecosystem mapping can be done in a hierarchy of classification units; types, communities, and species. The magnitude of the uncertainty decreases from the mapping unit of forest type to the mapping unit of forest community and to the mapping unit of forest species. This is because forest species are well defined, but the definitions of forest types are not clear-cut and are highly subjective. On the other hand, when an appropriate forest classification system and mapping unit (e.g., forest type in this study) has been chosen, data is required for the mapping process. This leads to the third source of the uncertainties in forest mapping, which is associated with the quantity and quality of the data.

We do not always have all the data we desire, and the data quality is not often satisfactory. For forest type mapping, the data may be incomplete, insufficient, and poorly qualified. Data incompleteness is due to lack of certain information in a specific place and at a specific time. For example, remotely sensed data for forest mapping may be incomplete in one part of a study area due to cloud cover or a technical problem with a sensor when passing the specific area.

Data insufficiency happens when one or more required data are not available so that the application does not proceed effectively. For example, remotely sensed data is often the only available data for forest type mapping. However, similar forest types present very similar responses in spectral space. Since spectral variation within forest types may be greater than between types, discrimination of forest types only on the basis of spectral reflectance is difficult, if not often impossible (Lees & Ritman, 1991; Moore *et al.*, 1991). Lees and Ritman (1991) have demonstrated that remote sensing is good at classifying coarse land cover classes in urban and rural/urban areas, i.e., Anderson *et al.* (1976) level I, but it performs poorly in densely forested areas. As a consequence, combining ancillary data with remotely sensed data becomes necessary in complicated forest type mapping. Moore *et al.* (1991) pointed out that there are associations between forest communities and the environmental variables that are correlated to their distribution; among these variables, elevation and geology have special significance through moisture and nutrient availability. Austin *et al.* (1983) showed that altitude along with its derivatives aspect and slope, and climatic factors, can be used to predict the distributions of several eucalypt species. This approach to integrating ancillary data with remotely sensed data is called "multisource" forest mapping.

Data quality, on the other hand, concerns the potential error originated from data measurement and pre-processing processes. All kinds of spatial data, whether they are paper maps or digital layers, in vector or raster format, having categorical or numerical values, contain errors to some extent (see Goodchild, 1989; Unwin, 1995). This is due to not only instrument and human errors, but also the age of the data and the inherent complexity of the real world. For example, there are some inherent limitations with remotely sensed data caused by significant atmospheric effects, topographic shadowing, and geometric errors, which can never be completely corrected (Richards, 1993). However, even if we have complete, sufficient, and good quality data for forest type mapping, applying spatial models or classifiers to handle this data leads to the fourth source of the uncertainties in forest mapping, which is associated with the imperfectness of predictive models.

There exist a large number of classifiers that have been used for forest mapping, none of which has been declared the best one. They all have limitations, and each of them may be suitable only for certain applications. For example, traditionally used parametric classifiers such as the Maximum Likelihood classifier are only good at classifying remotely sensed data into classes of Anderson *et al.* (1976) Level I (Lees & Ritman, 1991). It is very difficult to use the Maximum Likelihood Classifier to classify forest types at Anderson *et al.* (1976) Level III. More importantly, data error may be propagated and accumulated through the modelling process, which in most cases complicates the problem.

The above-mentioned four sources of the uncertainties in forest mapping are most often combined and pose great difficulties in the case of complicated forest type mapping such as this study. However, upon the selection of appropriate classifiers or predictive models, better classification performance can be obtained and more information can be conveyed to users and decision makers.

Traditionally, parametric or statistical classifiers such as the Maximum Likelihood Classifier, and the Minimum Distance classifier are used to classify remotely sensed data into forest classes (Chandrasekaran & Goel, 1988; Richards, 1993). Their performance has usually been poor (Skidmore and Turner, 1988). Civco (1993) describes the problems associated with these conventional, parametric-based approaches as being the production of fragmented, or salt-and-pepper-like, results; a lack of flexibility to permit variation in execution; an inability to integrate ancillary data; the lack of any means to consider spatial and contextual attributes; and, lastly, the inability to generalize or to incorporate potentially valuable non-statistical information. Nevertheless, studies (Hutchinson, 1982; Benediktsson *et al.*, 1990; Richards, 1993) have shown that traditional statistical methods can be used to integrate ancillary data into remote sensing classification. But this requires much greater effort and prior knowledge of data distributions. It is often difficult to implement these statistical methods for complicated real world problems. More importantly, these traditional algorithms do not cope well with noisy data so that the uncertainties associated with the data and the applications cannot be handled effectively.

Recently, however, it has been demonstrated that non-parametric models can overcome many of these limitations and show equivalent and sometime better performance, with less effort, on land cover classification. Among these models, Artificial Neural Networks, Decision Trees, Dempster-Shafer's Theory of Evidence, and Expert Systems are most prominent (e.g., Lee *et al.*, 1987; Skidmore and Turner, 1988; Skidmore, 1989; Hepner *et al.*, 1990; Benediktsson *et al.*, 1990; Lees and Ritman, 1991; Paola and

Schowengerdt, 1995; Fitzgerald and Lees, 1996). For example, Gahegan (2003) gave a strong boost for the role of machine learning methods such as Artificial Neural Networks and Decision Trees in dealing with complex geographical problems including land cover classification. But, he also conceded that much still needed to be done for wider use of machine learning methods.

In this study, a new strategy has been developed to classify multisource data into complicated forest types by using non-parametric models such as the Artificial Intelligence (AI) models mentioned above. The next section introduces AI models and their advantages and disadvantages in forest type mapping.

1.2 INDIVIDUAL AI MODELS AND THEIR ADVANTAGES AND DISADVANTAGES IN FOREST TYPE MAPPING

More than 30 years after the first Artificial Intelligence (AI) conference held at Dartmouth College in 1956, the definition of AI is still contentious among the insiders and outsiders of the field (Schank, 1988). Marvin Minsky defined Artificial Intelligence as "the science of making machines do things that would require intelligence if done by men" (Crevier, 1993). It implies that AI is closely associated with computer science and the science of cognition. Among the first wave of AI research in the early 1950s, General Problem Solver was a representative method developed by Hebert Simon and Allen Newell. The model was proved to be unsuccessful, as it intended to use a general model to solve problems that may have very diverse characteristics. However, it did inspire the development of two other AI models in the late 1950s and early 1960s. One is Frank Rosenblatt's Perceptron that is the first successful Artificial Neural Network. The other is Production System that later became an essential part of many Expert Systems. The first Expert System – DENDRAL came into the world in early 1970s developed by Edward Feigenbaum. This Expert System acquired knowledge from domain experts and represented knowledge as production rules (system). Around the same time, Decision Trees were developed in both the statistics and AI fields. These learn from examples and represent knowledge as decision trees. The revival of Artificial Neural Networks did not happen until the discovery of the multi-layer backpropagation neural network in 1986 (Rumelhart *et al.*, 1986).

In the past two decades, the AI models such as Expert Systems, Decision Trees, and Artificial Neural Networks have been successfully applied in many science and engineering fields including many AI applications in the geoscience field.

Depending on the knowledge acquisition methods, AI models can be divided into those that "learn from domain experts" and those that "learn from examples (machine learning)". On the other hand, depending on the knowledge representation methods, AI models can be divided into those that "represent knowledge as a symbol" and those that "represent knowledge as a sub-symbol". For example, traditional Expert Systems learn from domain experts and represent knowledge as symbols such as rules and frames. Decision Trees belong to a family of symbolic machine learning models. Artificial Neural Networks are considered as sub-symbolic machine learning models.

In this study, the AI models of Expert System, Decision Tree, Artificial Neural Network, and two other AI techniques, Dempster-Shafer's theory (Shafer, 1976) and fuzzy set theory (Zadah, 1965), are used. Both Dempster-Shafer's theory and fuzzy set theory are developments of the traditional Bayes theory for the purpose of handling uncertainty more effectively.

Compared with parametric classifiers such as the Maximum Likelihood Classifier, the Minimum Distance Classifier (MDC), and discriminant analysis, these AI models are non-parametric. One major drawback of the parametric models is that they either assume a particular statistical distribution that is usually not true in multisource data or require prior knowledge of data distributions (Benediktsston *et al.*, 1990). The non-parametric models, on the other hand, make no assumption about the distribution of the data. Moreover, the parametric models have to use complicated approaches such as those reviewed by Richards (1993), Benediktsston *et al.* (1990), and Hutchinson (1982) to combine ancillary information with remotely sensed data. The AI models, however, can readily handle multisource data without additional effort. Furthermore, the AI models have often been demonstrated to be better classifiers than parametric models in case of noisy data and complicated applications. Many applications reviewed in the next chapter have shown that the AI models could achieve generally higher classification accuracies than those of parametric models. In addition, the AI models can provide better classification understanding. This is particularly true when the mapping process can be captured as classification rules or decision trees, when inference logic such as fuzzy logic can be readily used in the modelling process and displayed through a user interface, and when uncertainties and confidence measures can be displayed as maps.

Unfortunately, there is no best AI model. They all have advantages and disadvantages in terms of computing requirement, classification performance, error tolerance, comprehensibility, and complexity. For example, the study of Shavlik *et al.* (1991) has compared a Decision Tree, ID3, and a backpropagation neural network. Their results showed that the backpropagation neural network performed slightly better than ID3, but it took much longer to train. The results also suggested that the backpropagation neural network handled noisy and incomplete data better than ID3. But obviously, classifications generated by ID3 were more easily interpreted than those of the backpropagation neural network, and they avoided the troublesome parameters and structure optimisation processes involved in the backpropagation neural network. This raises the research question of "can we improve classification performance and provide confidence measures by combining AI models?" The next section describes such a new strategy which uses combined AI models for complicated forest type mapping and attempts to improve classification performance and to provide information of classification confidence.

1.3 COMBINED AI MODELS AND THEIR ADVANTAGES IN FOREST TYPE MAPPING

The individual AI models used in this study are based on quite different principles. Therefore, they are likely to have different blind spots in mapping complicated forest types. These blind spots, however, can be cancelled out by combining the results of these individual AI models, which can potentially increase the predictive accuracy. Meanwhile, combining individual AI models can deliver one important element that individual AI models fail to deliver. It is believed that traditional accuracy assessment alone is not adequate to evaluate a classifier. An estimate of prediction confidence is also needed for every location (pixel) of the classified area. This is impossible to do using a single model without cross-reference in the predicted area. However, comparisons of classifications produced by different models can provide this element, especially when these models are based on different principles (Huang & Lees, 2004).

The idea of combining models is not new (e.g., Xu *et al.*, 1992; Rogova, 1994; Gahegan & West, 1998; See *et al.*, 1998). Several methods such as the majority voting system, Dempster's Rule of Combination, simple and weighted average methods, and production rule, have been used to combine the results of individual models. Studies have shown that combining results of individual models can be done either on an abstract level (i.e., hard classification output) or on a measurement level (i.e., soft classification output) (Xu *et al.*, 1992). In addition, the author will demonstrate that the combination can occur at several stages. For example, the results from the first combination stage can be again combined at the second stage to further improve classification performance.

In this study, three standalone AI models including a Decision Tree, an Artificial Neural Network, and a model based on Dempster-Shafer's theory were applied individually to the study area described in section 3.1. The hard and soft outputs of the three models were combined at the first and second stages using approaches applied in other studies and developed in this study. These include approaches based on the majority voting system and Dempster's Rule of Combination, approaches using simple statistical functions, and a group of weighted average approaches based on fuzzy set theory.

1.4 THE INDIVIDUAL AI MODELS AND THE COMBINED AI MODELS IN HANDLING DATA ERROR

As discussed above, error and uncertainty are ubiquitous in geographical applications, and data error can be propagated and accumulated through modelling processes due to the imperfectness of spatial models (see section 1.1). To date, a great deal of research has been dedicated to this topic. Veregin (1989) considered a 'hierarchy of needs' for modelling error in GIS operations as; error source identification, error detection and measurement, error propagation modelling, strategies for error management, and strategies for error reduction. For example, Crosetto and Tarantola (2001) suggested uncertainty and sensitivity analysis as tools for error detection and measurement and error propagation modelling. Agumya and Hunter (2002) on the other hand, recommended a risk management strategy for error management and reduction. This study also desired to identify apparent data error sources, evaluate the nature of the data errors, examine the propagation modes of the data errors through AI models, and recommend strategies for error management and reduction.

The three individual AI models used in this study all claim to be more error tolerant than traditional statistical models (e.g., Quinlan, 1986; Mingers, 1989; Hepner *et al.*, 1990; Moon, 1990). Is this true? How is data error propagated through these models? It is assumed that data error is propagated through the three models differently, as they are based on totally different principles. For Decision Trees, the relationships between independent variables and a dependent variable are captured as a hierarchical tree format. The independent variable that appears at the top level of the tree (i.e., root) obviously dominates others. Consequently, the data error associated with this variable will have the greatest negative effect on the classification results. Dempster-Shafer's theory, on the other hand, assumes independence among all independent variables, and no one variable is able to dominate others. The data error of one independent variable could be cancelled out by others. For a multi-layer feed-forward Artificial Neural Network, however, it is more complicated. The Artificial Neural Network is often called a 'black box' due to its complexity. The relationships between independent variables and a dependent variable are represented as the connection weights between input elements and hidden elements, and the connection weights between hidden elements and output elements. It is always very difficult to interpret these connection weights directly due to the hidden layers.

One might assume that, as the three individual AI models are likely to handle data error differently, by combining them the combined AI models might inherit the good performance of one model and suppress the bad performance of other models in dealing with the data error. This would potentially make the combined AI models more error tolerant.

After carefully examining the data sources, two apparent data errors have been identified in this study (see section 3.2 for details). One is sampling error, and the other is attribute error associated with the geology variable. Generally, to see how data error is propagated through a modelling process, it is necessary to utilize error models. Error models can be divided into two groups; formal mathematical models and simulation models. Formal mathematical models such as Taylor analysis (1982), MacDougall (1975) model, Newcomer and Szajgin (1984) model, Geman and Geman (1984) model, the Veregin (1989) model, Veregin (1995) method, and Goodchild *et al.* (1992) model, have been used to model error propagation through simple GIS overlay functions (e.g., RESELECT, AND, OR, XOR, Adding, Ratioing, Univariate overlay, logic functions, and area measurement) (Walsh *et al.*, 1987; Lanter & Veregin, 1992; Heuvelink *et al.*, 1989; Haining & Arbia, 1993; Arbia *et al.*, 1998; Drummond, 1987). Recently, simulation models such as Monte Carlo analysis have been strongly recommended for error propagation analysis (Openshaw, 1989; Lodwick, 1989). The advantage of simulation models over formal mathematical models is that their applications are not limited to simple GIS functions, but they are theoretically applicable to any function. For example, they have been used for the buffer function (Veregin, 1994; Veregin, 1996; De Genst *et al.*, 2001), DEM derivation (Lee *et al.*, 1992), logic model and continuous classification (Heuvelink & Burrough, 1993), and for Bayes theorem (Aspinall, 1992), as well as for GIS overlay functions (Openshaw *et al.*, 1991). Moreover, simulation models can help us gain insights into the modes of error propagation through the modelling process and lay a foundation to build a formal mathematical model (Veregin, 1994).

However, neither available formal mathematical models nor simulation models are applicable in this study. Two important assumptions for formal mathematical models are that only random errors influence the result (Drummond, 1987), and the modelling functions are continuously linearly differentiable (e.g., simple GIS overlay functions). This is clearly not the case for this study. Because the two identified data errors are systematic errors, and the three individual AI models are non-linear and non-differentiable, their principles are far too complicated to be simulated by mathematical functions. On the other hand, for simulation models, it is easy to peturb numerical data using an assumed error distribution model such as the normal distribution. But to corrupt categorical data, an exact probability distribution of any pixel belonging to different classes is required (e.g., Goodchild *et al.*, 1992). Among the seven independent input variables of this study, the geology variable is categorical data, but it has no known probability distributions for individual pixels (see section 3.2). This study, therefore, has developed methods to gain better insight into the modes of the three individual AI models in handling data errors and to evaluate the effectiveness of the combined AI models in handling the data errors through accuracy and visual assessments.

1.5 FUZZY EXPERT SYSTEMS AND FOREST MAPPING

This study has also developed several Fuzzy Expert Systems (Zadeh, 1983) for complicated forest type mapping. The major reason is that the three AI models and the combined AI models could not present comprehensible mapping to the end users. Expert Systems on the other hand, can interpret the mapping processes by representing knowledge as classification rules, and can capture the reasoning behind the mapping process through using formal logic.

The Fuzzy Expert Systems used a different strategy to deal with data error. While, the AI models mentioned above must assign a class to every pixel, the Fuzzy Expert Systems were set up to allow allocation of "unclassified" to pixels where the class allocation was not clear. This means that errors are not committed, and uncertainty is explicit.

The Fuzzy Expert Systems built in this study have the following components; a data base which can contain both fuzzy and crisp data, a knowledge base which is comprised of fuzzy production rules, an inference engine which is based on fuzzy logic, and a user-friendly interface that includes an explanation machine. One important difference between these Fuzzy Expert Systems and those of traditional Expert Systems is that instead of learning from domain experts, they learn classification rules directly from the samples that were selected either from the results of combined AI models or from the field survey. This could avoid the well-known knowledge acquisition "bottleneck" problems of the traditional Expert Systems (Grarratano & Riley, 1998; Mingers, 1986). In addition, representing the classification process as fuzzy production rules, and using fuzzy logic for the inference engine could facilitate the handling of the classification uncertainty.

1.6 OBJECTIVES OF THIS STUDY

Following discussion of the uncertainties and difficulties associated with forest type mapping, as well as the potential benefits of using individual AI models and combined

AI models for classification of multisource data and handling of data errors, the author has proposed the following objectives for this study and research.

- To compare the effectiveness of three individual AI models in mapping complicated forest types.
- To develop a new strategy for complicated forest type mapping by combining the three individual AI models.
- To examine the advantages of the combined AI models in terms of classification performance, confidence measures, time and resource requirement, and handling of the data errors.
- To gain insight into the modes of the three individual AI models in handling data errors.
- To develop Fuzzy Expert Systems for complicated forest type mapping which learn directly from samples.
- To examine the advantages of these Fuzzy Expert Systems compared with other individual AI models.

1.7 OUTLINE OF THIS STUDY

This study includes four stages. The first stage applies a Decision Tree, an Artificial Neural Network, and a model based on Dempster-Shafer's theory to multisource data and assesses the results for forest type mapping. The second stage applies a new strategy of combining the results of three AI models based on several combination methods in order to improve classification performance and provide confidence measures. The third stage applies methods to evaluate the modes of the three individual AI models and the effectiveness of the combined AI models in handling data errors. Fuzzy Expert Systems were then built in the last stage for forest type mapping by learning from the samples that were either derived from the results of combined AI models or selected from the field survey.

The layout of the thesis is as follows:

Chapter 1: Introduction

The chapter describes forest mapping and its uncertainties and difficulty, reviews past experiences of forest mapping, introduces the strategy developed in this study, and presents the potential advantages of this strategy.

Chapter 2: Literature review

In this chapter, the theory and applications of the models and techniques used in this study are reviewed in detail. These models and techniques are Artificial Neural Networks, Decision Trees, Dempster-Shafer's theory, fuzzy set theory, Expert Systems, and combination of models. For each of these models and techniques, the chapter first introduces its principles, then it reviews its applications in the land cover classification and geoscience fields. Following that it discusses and summarizes the advantages and disadvantages associated with each of these models and techniques.

Chapter 3: Study area, data and previous studies

The chapter describes the geographical location and natural environment of the study area. The input data including seven independent variables and dependent variable and their quality are discussed. The chapter then reviews the relevant previous studies at the study area in detail.

Chapter 4: Conceptual Model and Accuracy Assessment Methods

This chapter describes the conceptual model of this study. The conceptual model indicates that there were four modelling stages in this study, each of which focused on one topic. This chapter also describes the accuracy assessment methods used in this study for forest type mapping.

Chapter 5: Individual AI Models for Predictive Forest Type Mapping

The chapter presents the first stage of this study, in which three individual AI models including a Decision Tree, an Artificial Neural Network, and a model based on Dempster-Shafer's theory were applied to predictive forest type mapping using multisource data. First, the chapter describes the three individual AI models in detail. Then, the chapter reports the results of the three classifications in terms of predictive accuracies and visual appearance. Following that, the chapter discusses the results and findings. Finally, the chapter gives a short summary of this stage of study.

Chapter 6: Combined AI Models for Predictive Forest Type Mapping

The chapter presents the second stage of this study, in which the results of the three individual AI models were combined using different approaches for predictive forest type mapping. Firstly, the chapter describes the combined AI models in detail. Then, the chapter reports the results of the combined AI models for forest type mapping in terms of predictive accuracies and visual appearance. Following that, the chapter discusses the results and findings. Finally, the chapter gives a short summary of this stage of study.

Chapter 7: Evaluating the Modes of the Individual AI Model and the Effectiveness of the Combined AI models in Handling the Data Errors

The chapter presents the third stage of this study, in which the modes of the three individual AI models and the effectiveness of the combined AI models in handling data errors were examined. Firstly, the chapter describes the methods applied in this stage. Then, the chapter reports the results of the evaluations. Following that, the chapter discusses the results and findings. Finally, the chapter gives a short summary of this stage of study.

Chapter 8: Fuzzy (Rule-Based) Expert Systems for Predictive Forest Type Mapping

The chapter presents the fourth stage of this study, in which four Fuzzy (rule-based) Expert Systems were built from learning samples and applied to predictive forest type mapping. Firstly, the chapter describes the methods to build the Fuzzy Expert Systems in detail. Then, the chapter reports the results of the Fuzzy Expert Systems for forest type mapping. Following that, the chapter discusses the results and findings. Finally, the chapter gives a short summary of this stage of the study.

Chapter 9: Comparisons of Predictive Models

The chapter first describes the statistical Z-test method for comparing the classifiers of this study. Then, the chapter reports the results of the Z-test for the 34 individual and combined AI models applied in this study, and discusses the findings. Following that, the chapter compares the models of this study and those of previous studies. Finally, the chapter gives a short summary for the chapter.

Chapter 10: Conclusions

The conclusions of this study based on the results and the discussions are given in the chapter. The chapter presents the benefits and drawbacks of each of the three individual AI models, summarizes the advantages of combination strategy, and concludes with a discussion of the usefulness and limitations of Fuzzy Expert Systems in mapping complicated forest types and handling data errors.

1.8 PUBLICATIONS FROM THIS STUDY

1. Huang, Z., 2001, Combined Non-Parametric based models for multisource predictive forest mapping, *Proceedings of the 6th International Conference on GeoComputation*, Sep 2001, Brisbane.

2. Huang, Z., and Lees, B.G., 2002, Combining AI models in dealing with input errors and improving classification performance, *5th International Symposium on Spatial Accuracy Assessment in Natural Resource and Environmental Sciences*, July 2002, Melbourne, Edited by Hunter, G., and Lowell, K., pp 512-521.

3. Huang, Z., and Lees, B. G., 2004, Combining Non-Parametric Models for Multisource Predictive Forest Mapping, *Photogrammetric Engineering and Remote Sensing*, Vol 70, 415-426.

Chapter 2

LITERATURE REVIEW

In this chapter, the theory and application of the models and techniques used in this study are reviewed in detail. These models and techniques are Artificial Neural Network, Decision Tree, Dempster-Shafer's theory, Fuzzy Set Theory, Expert System, and combination of models. For each of these models and techniques, the chapter first introduces its principles, then it reviews its applications in the land cover classification and geoscience fields, after that it discusses and summarizes the advantages and disadvantages associated with each of these models and techniques.

2.1 ARTIFICIAL NEURAL NETWORK

Artificial Neural Networks have attracted a large number of applications in the field of classification in the past decade, because they are useful in problems which violate the fundamental assumptions of parametric methods and can provide solutions to problems previously considered intractable (Lees, 1996b). They cope better than what with incomplete and noisy data, and produce approximate results through parallel learning (Obermeier & Barron, 1989). Generally speaking, what makes Artificial Neural Networks so attractive is their capability to deal with problems that are too complicated to set up using traditional methods.

2.1.1 Principles of Artificial Neural Networks

The power of Artificial Neural Networks relies on their capacity to act as universal approximators (Hornik, 1989). For example, a standard multi-layer feed-forward Artificial Neural Network with as few as one hidden layer using arbitrary activation (squashing) functions is capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficiently many hidden units are available (Hornik, 1989). The general idea of Artificial Neural Networks is the adaptation of network weights during the learning process in order to approximate functions or to discriminate classifications (Gahegan & West, 1998). For example, for a three-layer feed-forward Artificial Neural Network, each hidden neuron along with its associated weights to the input layer represents a linear hyperplane within feature space (German *et al.*, 1997). Adjustment of the weights means moving these hyperplanes to better discriminate feature space. Complex boundaries can be built by combining these hyperplanes in order to separate non-linear feature space. This is made possible by adjusting the weights between the hidden and output layers.

Six popular Artificial Neural Network algorithms have been reviewed by Lippmann (1987), and they are the Hopfield net, the Hamming net, the Carpenter/Grossberg net, Perceptron, multi-layer Perceptron, and Kohonen's Self Organizing Maps. Carpenter (1989) provides a review of several Artificial Neural Networks that can be used for pattern recognition. They are the McCulloch-Pitts neuron, Perceptron, Adaline and Madline, Backpropagation, learning matrix, Linear Associative Memory, embedding fields, Instars and Outstars, avalanche, Shunting Competitive Networks, competitive learning, computational mapping, Adaptive Resonance Theory, Cognitron and Neocognitron. Among these Artificial Neural Networks, the Artificial Neural Network using a backpropagation algorithm is the most popular, especially in the land cover classification field; therefore it deserves a detailed introduction.

The standard backpropagation training algorithm uses an iterative gradient method to minimise the error function between the actual output and the desired output. There are five steps associated with the algorithm (Lippmann, 1987):

1. Initialise weights and offsets
2. Present new input and desired output
3. Calculate actual output
4. Adapt weights to minimise the error function
5. Repeat by going to step 2

Below is a simple description of the standard backpropagation algorithm (Rumelhart *et al.*, 1986). Without losing generalization, a three-layer feed-forward Artificial Neural Network is built as follows (Figure 2.1):

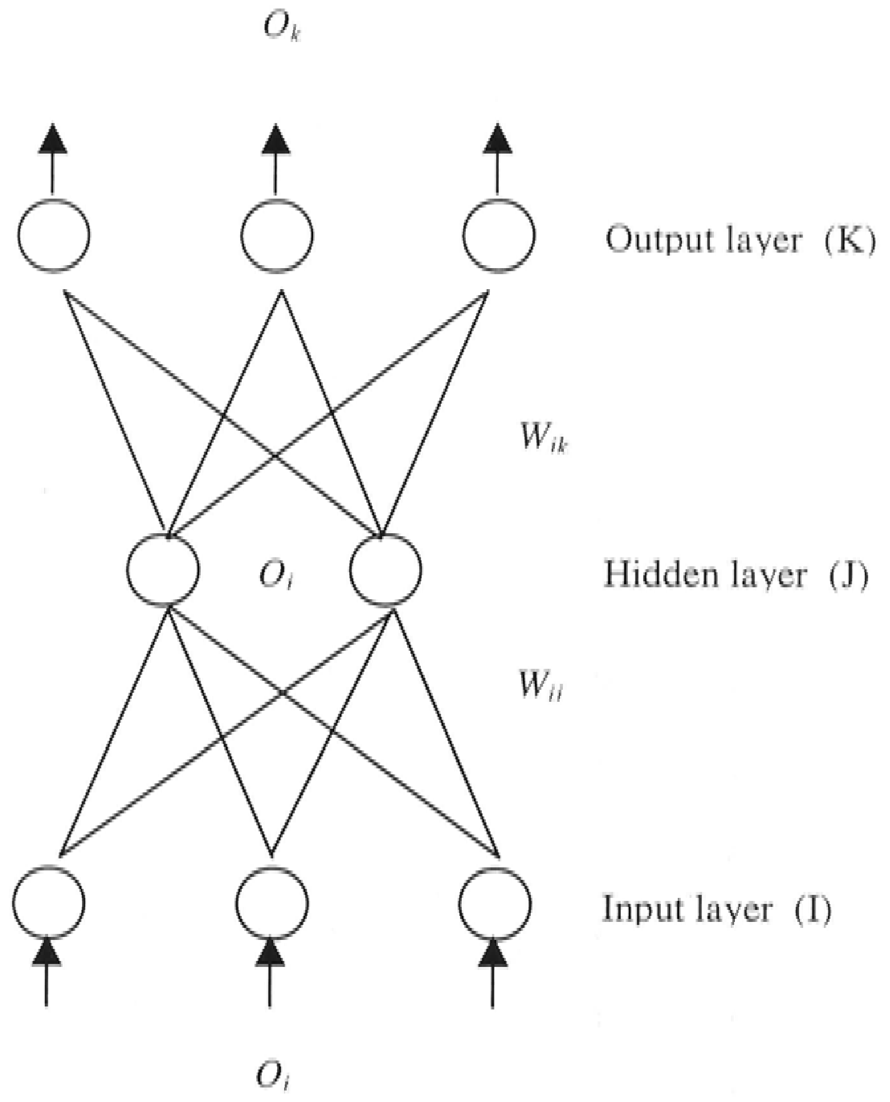


Figure 2.1 An example of three-layer feed-forward neural network

Where W_{ij} is the weight between the i_{th} input layer and the j_{th} hidden layer, W_{jk} is the weight between the j_{th} hidden layer and the k_{th} output layer, O_i is the i_{th} pattern input, O_j is the output of the j_{th} neuron on the hidden layer, and O_k is the output of the k_{th} neuron on the output layer. We assume that the activation function of the Artificial Neural Network is a *sigmoid logistic function*, and the error function of the Artificial Neural Network is a *Summed Squared Error (SSE) function*.

Feed-forward stage

(1) Input of the j_{th} hidden layer neuron:

$$Z_j = \sum_i W_{ij} \times O_i \tag{2.1}$$

(2) Output of the j_{th} hidden layer neuron:

$$O_j = \frac{1}{1+\exp(-Z_j)} \tag{2.2}$$

(3) Input of the k_{th} output layer neuron:

$$Z_k = \sum_j W_{jk} \times O_j \tag{2.3}$$

(4) Output of the k_{th} output layer neuron:

$$O_k = \frac{1}{1+\exp(-Z_k)} \tag{2.4}$$

The total error, E is defined as:

$$E = \frac{1}{2}\sum_{i}\sum_{k}(T_{ik} - O_{ik})^2 \qquad (2.5)$$

Where T_{ik} is the *desired output*, and O_{ik} is the *actual output* of the k_{th} output layer neuron. A backpropagation stage then minimizes E by using the gradient descent method or the *generalized delta rule*.

Backpropagation stage

(5) Error of k_{th} output layer neuron:

$$E_k = \frac{\partial E}{\partial O_k} = (T_k - O_k)(1 - O_k)O_k \qquad (2.6)$$

(6) Weight adjustment of W_{jk} is:

$$\Delta W_{jk} = \eta E_k O_j \qquad (2.7)$$

where η is the *learning rate*

(7) Error of the j_{th} hidden layer neuron:

$$E_j = \frac{\partial E}{\partial O_j} = O_j(1 - O_j)\sum_{k} E_k W_{jk} \qquad (2.8)$$

(8) Weight adjustment of W_{ij} is:

$$\Delta W_{ij} = \eta E_j O_i \qquad (2.9)$$

To speed up convergence, and to avoid being trapped in local minima in the error surface, a *momentum term* (α) can often be added to the equations, i.e.

(9) $\Delta W_{jk}(t) = \eta E_k O_j + \alpha\Delta W_{jk}(t\text{-}1)$ (2.10)

(10) $\Delta W_{ij}(t) = \eta E_j O_i + \alpha\Delta W_{ij}(t\text{-}1)$ (2.11)

where t is the current epoch number.

2.1.2 Applications of Artificial Neural Networks

There are a large number of applications of Artificial Neural Networks since the 1980s. For example, Atkinson and Tatnall (1997) conducted a review of the use of Artificial Neural Networks in remote sensing applications. What we are concerned about, however, is the field of land cover classification problem using Remote Sensing and GIS data, therefore only applications belonging to this topic are reviewed next.

Researchers in the Remote Sensing field were among the first group to adopt Artificial Neural Networks as an alternative classifier to the Maximum Likelihood Classifier. This is mainly due to well-known difficulties associated with the parametric-based Maximum Likelihood Classifier in dealing with high dimensional remotely sensed data and in combining ancillary information. Artificial Neural Networks, on the other hand, are non-parametric, and have advantages over the Maximum Likelihood Classifier in dealing with incomplete data, noisy data, and multisource data. More importantly, they are able to delineate non-linear feature space that occurs occasionally in land cover classifications.

The first group of Artificial Neural Network applications uses multi-layer feed-forward standard backpropagation Artificial Neural Networks for the classification of Remote Sensing imagery. The results are often compared to parametric methods in order to get a sense of the advantages and problems of this Artificial Neural Network. In many cases, the Artificial Neural Network could achieve better or comparable overall classification accuracies compared with conventional parametric models such as the Maximum Likelihood Classifier and discriminant analysis, especially in high variance classes (e.g., Key *et al.*, 1989; Ritter & Hepner, 1990; Heermann & Khazenie, 1992; Hara *et al.*, 1994; Yoshida & Omatu, 1994; Paola & Schowengerdt, 1995; Jarvis & Stuart, 1996; Foschi & Smith, 1997; Bischof *et al.*, 1992), but it may also be inferior to the Maximum Likelihood Classifier (Civco, 1993). The advantage of using the standard backpropagation Artificial Neural Network to classify remotely sensed data is not obvious, as remotely sensed data is parametric data which fits in with the assumption made by conventional parametric models such as the Maximum Likelihood Classifier.

In case of complicated land cover classification, ancillary information such as textural information and other environmental data is often required to be used along with the remotely sensed data to help improve the classification. This is usually called "multisource" classification. Artificial Neural Networks have obvious advantages in classifying multisource data over the parametric methods. Firstly, multisource data often have many data types, some of which may violate the assumption of Gaussian distribution made by the Maximum Likelihood Classifier. However, Artificial Neural Networks are non-parametric models which do not assume a statistical distribution. Secondly, Artificial Neural Networks can automatically combine multisource data and determine the weights between data sources without any difficulty. However, statistical methods such as the Maximum Likelihood Classifier have to use complicated approaches and require prior knowledge of data distributions to combine ancillary information (e.g., Richards, 1993; Benediktsson *et al.*, 1990; Hutchinson, 1982). Several studies have been done using the standard backpropagation Artificial Neural

Network for classification of multisource data, and the results were better or comparable with less effort compared with the Maximum Likelihood Classifier (e.g., Hepner *et al.*, 1990; Benediktsson *et al.*, 1990; Zhuang *et al.*, 1991; Bischof *et al.*, 1992; Gong *et al.*, 1996; Gong, 1996; Karminsky *et al.*, 1997). Nevertheless, several problems associated with the standard backpropagation algorithm have also been identified, these include slow convergence on a solution, local minima, sample representation, overtraining, and computational complexity.

The most serious problems of the standard backpropagation feed-forward multi-layer Artificial Neural Networks are that they are relatively slow to converge on a solution, and that they can be trapped in local minima. Therefore, some modifications of the standard backpropagation algorithm and some optimisation methods have been recommended for the multi-layer Artificial Neural Network to deal with these problems (Alpsan *et al.*, 1995). One such recommendation is to use an optimisation algorithm to remove insignificant hidden neurons, which could result in improved classification performance (Dreyer, 1993). Other modified backpropagation algorithms used for the geoscience applications include the Blocked Back-Propagation (BBP) algorithm developed by Liu and Xiao (1991), the Quickprop algorithm that executes much faster in training (Foody, 1995; Foody, 1996; Foody *et al.*, 1995), the conjugate-gradient algorithm (Benediktsson *et al.*, 1993; Benediktsson & Sveinsson, 1997; German & Gahegan, 1996), and the Quasi-Newton optimisation method (Fischer & Staufer, 1999). Generally speaking, these modified backpropagation algorithms converge much faster and are able to improve classification performance. However they still cannot guarantee an optimal result, and sometimes they still require extra time to determine the network architecture and parameters.

In recent years, multi-layer feed-forward Artificial Neural Networks using algorithms other than the backpropagation have received attention in the field of land cover classification. For example, one such Artificial Neural Network is the Binary Diamond Neural Network applied by Salu and Tilton (1993) and Murnion (1996). The two studies showed that the Binary Diamond Neural Network algorithm is easier to use, and it can be trained more rapidly than the standard backpropagation Artificial Neural Network. The classification performance of the Binary Diamond Neural Network was better than the standard backpropagation Artificial Neural Network in the case study of Salu and Tilton (1993), but it was worse in the case study of Murnion (1996).

Meanwhile, Artificial Neural Networks other than the multi-layer feed-forward network also began to emerge from the applications of land cover classification. For example, Tzeng *et al.* (1994), Chen *et al.* (1995), and Chen *et al.* (1997) have suggested a dynamic learning (DL) Artificial Neural Network based on a Kalman filtering technique for classifying remotely sensed data. Their studies have indicated that the DL Artificial Neural Network has advantages of fast convergence, a built-in optimisation function, and global scale. Moreover, the Learning Vector Quantization (LVQ) Artificial Neural Network and its extended versions have been reported to have achieved better classification performance than the backpropagation Artificial Neural Network in some cases (e.g. Hara *et al.*, 1994; Ito & Omatu, 1999; Corne *et al.*, 2000). In addition, applications using another popular Artificial Neural Network – ART and its family have also been widely reported (e.g., Carpenter *et al.*, 1992; Hara *et al.*, 1994; Carpenter *et al.*, 1997; Carpenter *et al.*, 1999; Carpenter *et al.*, 1999; Gopal *et al.*, 1999; Pax-Lenney *et al.*, 2001). The classification performances of these studies, however, were not consistent. The Kohonen self-organising map and Hopfield neural network can also be used for classification (Openshaw & Turton, 1996; Tatem *et al.*, 2002). However, none of these Artificial Neural Networks using other than the standard backpropagation algorithms has been demonstrated to be the best overall for the land cover classification problem.

2.1.3 Problems of Artificial Neural Networks

Even though Artificial Neural Networks can achieve generally better results than conventional statistical methods, there are quite a few factors that can influence the performance of Artificial Neural Networks, especially the standard backpropagation Artificial Neural Network. Some studies have been done to examine the sensitivity of the factors that may affect the classification performance (Jarvis & Stuart, 1996; Foody & Arora, 1997). Factors such as learning algorithms, sampling sizes, network architectures, parameter selections, and activation functions are now described, and research on these issues is reviewed below.

Foody and Arora (1997) reported an evaluation of four factors that can affect the classification accuracy of backpropagation Artificial Neural Network. These factors are the dimensionality of the remotely sensed data, the network architecture, and the characteristics of the training and test sets. The results revealed that the network architecture did not affect the classification accuracy significantly, which has also been confirmed by Jarvis and Stuart (1996). However, the other three factors did have a large effect on the classification accuracy. For example, the classification accuracy was positively increased with the size of the training set, the number of the remote sensing wavebands, and the size of the test set. On the other hand, their study showed that increase of these factors would raise the complexity of networks and eventually require more training time and computing resources, which may not be compensated for by an increase in the classification accuracy.

Meanwhile, the choice of an appropriate error (criterion) function is another very important issue in the design of an Artificial Neural Network. It is to optimize the error function that we can expect an optimal performance of a classification. Four such criterion functions were compared by Barnard and Casasent (1989). They are a Least Mean Squares (LMS) function, a Perceptron function, a Bayesian function, and a sigmoid function. Two issues were considered important in selecting a criterion function; the error rate and the existence of local minima. The authors found that the sigmoid function was the best overall, followed by the Perceptron, the Bayesian and the LMS functions. They also pointed out that the sigmoid function used by the backpropagation Artificial Neural Network was the reason behind the better performance of this kind of Artificial Neural Network over conventional linear classifiers. A case study comparing an adaptive-clustering classifier with a Gaussian classifier on an artificial data set confirmed this conclusion.

How to optimise the size of training samples is yet another problem, especially when using the backpropagation Artificial Neural Network for classification. A study by Zhuang *et al.* (1994) used the backpropagation Artificial Neural Network to classify a Landsat TM dataset. The training sample sizes were varied from 5%, 10%, 15%, to 20% of the TM data. The results showed that there were no differences among the classifiers using 5%, 10%, and 15% samples, and between the classifiers using 15% and 20% samples, but the classifiers using 5% and 10% samples did differ from the classifier using 20% samples. They concluded that approximately 5-10% of the image data was needed to train an Artificial Neural Network adequately to obtain satisfactory performance, however, further increasing the training sample sizes did not necessarily improve the classification accuracies, but did greatly increase the training time.

Meanwhile, the findings of Gopal *et al.* (1999) showed that there may be a minimum size of training sample required for each class in order to learn the class' characteristics effectively, and it is the quality (representation) of the training data that had more effect on the classification performance. On the other hand, Blamire (1996) assessed the influence of relative sample size in training Artificial Neural Networks. A backpropagation Artificial Neural Network was used to classify a Landsat TM dataset into just two classes; built and soil. As the built class is more complex, the author expected that the soil class would be over-represented if identical training samples for both classes were used. Therefore, the study fixed the training sample size to 90 pixels for the built class and decreased the training sample sizes of the soil class to 60, 45, and 30 pixels. It was found that there was relatively little variation in the overall accuracies with the decrease of the relative sample sizes of the soil class except with the case of 30 pixels. Thus, Blamire suggested that relative sample sizes for training should be appropriately selected. Generally speaking, Artificial Neural Networks need fewer training samples than the conventional statistical models and Decision Trees (Hepner *et al.*, 1990; Lees, 1996c), because they use training samples more efficiently through parallel learning instead of sequential learning. But they may be more sensitive to sample representation and may require a minimum number of samples for each class. Also, if an Artificial Neural Network is used to classify multisoure data, probably only about 0.5-1% of the classified area is needed as samples to train the Artificial Neural Network adequately to obtain satisfactory classification performance.

One of the most serious problems of the backpropagation algorithm, however, is that a local minimum instead of a global minimum may be found during the network implementation. Because the standard backpropagation algorithm as well as the conjugate-gradient algorithm belongs to the family of local optimisation methods, when the error surface is flat, when the gradients are in a large range, and when the error surface is very rugged, they are easily caught in a local minimum (Rumelhart *et al.*, 1986). Global search algorithms, on the other hand, avoid this by using such approaches as restarting from new initial weights each time and using the simulated annealing algorithm. Evolutionary methods and genetic algorithms are other global optimisation methods. For example, Fischer *et al.* (1999) recommended using the Differential Evolution Method for the purpose of avoiding the parameter selection problem of Artificial Neural Networks. The approach is a global optimisation algorithm that employs a structured, randomised parallel multipoint search strategy to push the search out of local minima when the error function has relatively low values. The results of their case study have indicated good performance on both training and testing. On another study, a genetic algorithm was used to evaluate the advantages of global search strategies over the standard backpropagation algorithm (Sexton and Dorsey, 2000). Ten datasets were used for the classifications. The results showed that the genetic algorithm achieved better classification accuracies than the standard backpropagation Artificial Neural Network for all 10 datasets. There were no consistent results for computation time between the standard backpropagation Artificial Neural Network and the genetic algorithm. The genetic algorithm has also achieved overall better performance for the testing sets. They pointed out that the genetic algorithm was more consistent and predictable than the standard backpropagation Artificial Neural Network due to the independence of its initial random weights.

Another problem of Artificial Neural Networks is the so called 'black box" problem. This is because the network behaviour is very difficult to interpret. However, research has been undertaken to gain a better insight into how an Artificial Neural Network does work by using approaches such as WV-Curves (Bischof *et al.*, 1992), Casual-Index (Enbustu *et al.*, 1993), GIS visualisation (Laffan, 1998), network prunings (Abrahart *et al.*, 1998), and tree-like network (Serpico & Roli, 1995).

Looking at all of these problems of Artificial Neural Networks, Kanellopoulos and Wilkinson (1997) explored several strategies and identified the best practice for neural network classification. They recommend;

1. Use normalized input instead of raw data,
2. Change the learning rate and momentum to prevent chaos effects in which small changes of inputs will cause very large changes in output,
3. Apply fast convergence learning algorithms such as the conjugate-gradient method,
4. Experimentally determine the network architecture,
5. Use ancillary information, and
6. Use combined or hybrid classifiers.

Among these strategies, using combined or hybrid classifiers can give us several notable advantages. It can avoid several problems associated with standard backpropagation Artificial Neural Network. For example, the Artificial Neural Network fed into a combined classifier does not need to spend extra time in optimising the network architecture and parameters.

2.1.4 Summary of Artificial Neural Networks

The above review reveals that there have been a large number of applications of Artificial Neural Networks in the land cover classification field. This is because Artificial Neural Networks have demonstrated generally better overall performance than conventional statistical methods, especially when multisource, non-linear classifications are involved. It should be noted that if conventional statistical methods are properly set up for multisource classification, they can achieve better or comparable performance than backpropagation Artificial Neural Network; but this requires much greater effort and prior knowledge of data distributions (Benediktsson *et al.*, 1990).

Several factors have been shown to have significant effects on network performance and increase the complexity of Artificial Neural Networks. These factors include network architectures, parameter selections, initial weights, sample sizes, sample representation, data representation, activation functions, and error functions, etc. The standard backpropagation Artificial Neural Network was found to be suffering from relatively slow convergence, potential of local minima, and instability of performance. To ease the problems, powerful modern computer hardware, algorithms such as the one suggested by German & Gahegan (1996) that can automatically select network architecture, parameters and weights, faster learning algorithms such as the conjugate-gradient approach, global search strategies such as the genetic algorithms and the Differential Evolution Method, other neural networks such as the LVQ, and the combined/hybrid classifiers have been suggested.

2.2 DECISION TREES

Decision Trees are another group of machine learning algorithms for decision-making and pattern recognition based on inductive learning from samples. They are robust methods for classification problems, especially when multisource data and non-linear feature spaces are involved. They enjoy advantages such as noise and uncertainty tolerance, computational efficiency, and tree-like symbolic production which is comprehensible.

2.2.1 Principles of Decision Trees

Decision Trees are able to produce hyperplanes like Artificial Neural Networks for discriminating classes, only not through adjusting weights, but through the path described from the root of a Decision Tree to a leaf (Gahegan and West, 1998). Generally, there are three phases involved in constructing a Decision Tree for a classification problem; building a tree structure from samples, pruning initial trees to improve effectiveness, and applying the pruned tree for classification (Mingers, 1989a).

Four heuristic methods have been used to construct Decision Trees (Safavian & Landgrebe, 1991). They are Top-Down approaches, Bottom-Up approaches (e.g., Landeweerd *et al.* 1983), Hybrid approaches, and tree Growing-Pruning approaches. Most of the popular Decision Trees such as CART (Classification And Regression Tree) (Breiman *et al.*, 1984), ID3 (Quinlan, 1986), CN2 (Clark & Niblett, 1989), and C4.5 (Quinlan, 1993) use the Top-Down approaches. Therefore they rely on measures of "goodness of split" (splitting rules) to select the most appropriate features (input variables or attributes) on each internal node of a Decision Tree to split the branches. These measures of "goodness of split" can be divided into two large groups (Quinlan, 1990). The first group is based on information or entropy theory (e.g., Hartmann *et al.*, 1982; Casey & Nagy, 1984; Goodman & Smyth, 1988; Quinlan & Rivest, 1989), which originated in the AI field. The second group is based on statistical theory (e.g., Friedman, 1977; Rounds, 1980; Qing-Yun & Fu, 1983; Schuermann & Doster, 1984; Li & Dubes, 1986), which originated in the statistics field. One popular measure of "goodness of split" from the first group is the information gain of Quinlan (1986), while two examples of the second group are the *GINI* measure (Breiman *et al.*, 1984) and the X^2 (Hart, 1985).

Mingers (1989a) has reviewed, and experimentally compared, six measures of "goodness of split" from the above two groups. They are the information measure (*IM*), the chi-square statistic (X^2), the G statistic, the probabilities measure (*PROB*), the *GINI* index of diversity, the Gain-ratio measure (*GR*), and the Marshall correction (*MARSH*). These six splitting rules are reviewed as follows.

The IM evaluates the information gain of each internal node, and chooses the attribute to branch on which it gains the most information (Quinlan, 1986). For example,

$$IM = \frac{1}{N}\left(\sum_i \sum_j X_{ij} \log(X_{ij}) - \sum_i X_{i.} \log(X_{i.}) - \sum_j X_{.j} \log(X_{.j}) + N \log(N)\right) \tag{2.12}$$

where IM is the information measure, N is the total number of samples, X_{ij} is the number of the i_{th} value of evaluated attribute which be classified as the j_{th} class, $X_{i\cdot}$ is the row sum of the i_{th} value, and $X_{\cdot j}$ is the column sum of the j_{th} class in a contingency table or confusion (error) matrix. This algorithm based on mutual information theory is near optimal (Goodman & Smyth, 1990), but it tends to favor attributes with many values (Quinlan, 1986).

The X^2 statistic measures the association between two variables in a contingency table, and chooses the attribute with greater X^2 value. For example,

$$X^2 = \frac{\sum_i \sum_j (X_{ij} - E_{ij})^2}{E_{ij}} \tag{2.13}$$

where $E_{ij} = \dfrac{X_{i.} X_{.j}}{N}$ is the expected value for the i,jth cell in the error matrix.

The G statistic is another statistic based on information theory. For example,

$$G = 2N \times IM \tag{2.14}$$

where IM is the information measure calculated from equation 2.12.

The $PROB$ calculates the probability of an attribute value occurring from the X^2 distribution or the G statistic. Attribute with smaller probability is chosen for the node.

The $GINI$ function measures the "impurity" of an attribute with respect to the classes (Breiman *et al.*, 1984). Then it chooses a split that minimizes the "impurity", for example:

$$I(t) = 1 - \sum_j \left(\frac{X_{.j}}{N}\right)^2 \tag{2.15}$$

If $I(t)$s are equal for two attributes, then another measure must be used:

$$I = \frac{1}{N}\left(\sum_i \sum_j \frac{X_{ij}^{\ 2}}{X_{i.}} - \sum_j \frac{X_{.j}^{\ 2}}{N}\right) \tag{2.16}$$

The attribute with larger value of I is chosen for branching the node.

The GR measure proposed by Quinlan (1986) is:

$$GR = \frac{IM}{IV} \tag{2.17}$$

where $IV = \dfrac{-\sum_i X_{i.}}{N \log(\frac{X_{i.}}{N})}$ is the information value and IM is the information measure calculated from equation 2.12. The GR measure favors those attributes with an unequal distribution of samples and with a small number of values.

The $MARSH$, on the other hand, is used to avoid producing small splits, and favors attributes with an even distribution of samples. For example,

$$MARSH = A \times \frac{X_{1.}}{N} \times \frac{X_{2.}}{N} \times \cdots \times k^k \tag{2.18}$$

where A is any measures of "goodness of split", k is the number of attribute values.

In order to evaluate the effectiveness of these splitting rules, they were applied to four datasets, and the sizes of the Decision Trees and the accuracies of the Decision Tree classifications were compared (Mingers, 1989a). The results showed that the GR measure produced the smallest tree, and the X^2 has produced the largest, while the $GINI$ and the G statistic were in between. Meanwhile, it was found that using the $PROB$ measure and the $MARSH$ correction increased the size of trees. The accuracy results, however, did not reveal any difference among these splitting rules. In addition, when a pruning algorithm was implemented, the sizes of trees were significantly reduced, and the accuracies were improved. Mingers (1989a) concluded that the choice of splitting rules did not significantly influence the sizes and the accuracies of the pruned trees, and that a random splitting rule did not significantly decrease classification accuracy. However, further studies of Buntine and Niblett (1992) and Liu and Whilte (1994) showed that the random splitting rule did lead to a significant decrease in classification accuracy.

Besides the above reviewed six measures of "goodness of split", other measures of "goodness of split" have also been used, such as the subset criterion of ASSISTANT

(Kononenko *et al.*, 1984). Safavian and Landgrebe (1991) have provided a detailed review of many of these measures.

Initial Decision Trees created by using the measures of "goodness of split" are usually very large, and therefore they suffer from inefficiency and lack of comprehensibility. As stated by Quinlan (1986), a simple tree is preferred because of its higher likelihood to capture structure inherent in the problem. Thus, pruning algorithms are always used to remove the branches and nodes that give little information.

Quinlan (1987) has reviewed three pruning algorithms for Decision Trees, which include the Cost-Complexity pruning (Breiman *et al.,* 1984), the Reduced Error pruning, and the Pessimistic Pruning. His results showed that the Cost-Complexity algorithm produced smaller trees than both of the Reduced Error and the Pessimistic algorithms. Meanwhile, the pruned Decision Trees also achieved better or equivalent accuracies than the initial Decision Trees.

Later, Mingers (1989b) experimentally compared five pruning algorithms of Decision Trees, three of which have been reviewed by Quinlan (1987). Descriptions of these pruning algorithms are as follows.

The Cost-Complexity method was initially developed by Breiman *et al.* (1984). It takes into account both the misclassification errors and the complexity (size) of the tree. The idea is to equal the Cost-Complexity measures before and after pruning the sub-tree of a certain node. For example, before the sub-tree of the node is pruned, the Cost-Complexity for the sub-tree of the node is:

$$\frac{M_T}{N} + \alpha N_T \tag{2.19}$$

While after the sub-tree of the node is pruned, the Cost-Complexity of the node is:

$$\frac{M_t}{N} + \alpha \tag{2.20}$$

where M_T is the number of test samples misclassified before pruning, M_t is the number of test samples misclassified after pruning, N is the total test samples, N_T is the number of leaves under the node considered, and α is a parameter. If we assume that

$$\frac{M_T}{N} + \alpha N_T = \frac{M_t}{N} + \alpha \tag{2.21}$$

Then $\alpha = \dfrac{M_t - M_T}{N \times (N_T - 1)}$ (2.22)

The algorithm calculates α for each internal node (except the root), and select the nodes with the smallest α to prune. The algorithm, however, needs a separated test set.

The Critical Value pruning method estimates the importance or strength of a node from calculations done in the initial tree creation stage (Mingers, 1987a). The approach specifies a critical value from the measures of "goodness of split", and prunes those nodes that do not reach the critical value along the branch.

Niblett and Bratko (1987) have described the Minimum-Error pruning method to find the single tree that could give the minimum error rate when classifying a test set. The expected error rate E_k is given by:

$$E_k = \frac{n - n_c + k - 1}{n + k} \quad (2.23)$$

where n is the number of samples on the node, n_c is the number of samples correctly classified as class c, and k is the total number of classes. The sub-tree of the node will be pruned if the expected error rate after pruning is smaller than that before pruning. The algorithm does not require a separated test set, but it suffers from drawbacks such as an assumption of equally likely classes (which is often not satisfied), and production of only a single tree. It was also found that the number of classes (i.e., the k) has significant effects on the final outcomes.

The Reduced-Error method recommended by Quinlan (1987) can be divided into several implementation stages. It first uses an initial tree to classify a test set. For each node, it then counts the number of samples wrongly classified if the sub-tree is kept and if it is removed. Next, the difference between them (if positive) is measured as the gain from pruning the sub-tree. Subsequently, it chooses the sub-tree with the largest of gains to be actually pruned.

The Pessimistic-Error algorithm (Quinlan, 1987) calculates error rates as:

$$E_t = M_t + \frac{1}{2} \quad (2.24)$$

for a node, and

$$E_T = M_T + \frac{N_T}{2} \qquad (2.25)$$

for the sub-tree of the node. Standard error of the sub-tree is calculated as:

$$SE_T = \sqrt{\frac{E_T \times (N_T - E_T)}{N_T}} \qquad (2.26)$$

The sub-tree will be pruned if $SE_T + E_T >= E_t$. The algorithm does not need an independent test set, and it is also much faster than other methods.

To compare the effectiveness of the five pruning methods, they were applied to five datasets (Mingers, 1989b), and they were evaluated on the criteria of the tree sizes and the classification accuracies. The results showed that the five pruning methods did significantly reduce the size of the trees. Among them, the Cost-Complexity and the Critical Value methods produced the smallest trees, while the Minimum-Error algorithm produced the largest. For the accuracy measures, the Minimum-Error and the Pessimistic pruning were the least accurate, while the Reduced-Error and the Error-Complexity were the most accurate. It was also found that the pruning could greatly improve the classification accuracy under noisy circumstances.

There were experiments indicating that Decision Trees are able to cope with noisy data (i.e. conflicting data) (e.g., Quinlan, 1986; Mingers, 1989a; Liu & White, 1994). These experiments confirmed that low levels of noise does not cause the classification accuracy to reduce rapidly. Moreover, several methods have also been suggested for Decision Trees to learn effectively under a situation of incomplete data (Friedman, 1977; Breiman *et al.*, 1984; Quinlan, 1986). Safavian and Landgrede (1991) have provided a comprehensive survey of Decision Trees on all these relevant issues.

2.2.2 Applications of Decision Trees

There have been many applications of Decision Trees in the pattern recognition and classification fields during the past three decades. For example, Decision Trees have been used in speech analysis (Dattatreya & Sarma, 1981), edge detection (Goodman & Smyth, 1990), Chinese character recognition (Gu *et al.*, 1983; Wang & Suen, 1987), cervical cell classification (Qing-Yun & Fu, 1983), white blood cell classification (Landeweerd *et al.*, 1983), and other pattern recognition and classification problems (e.g., Li & Dubes, 1986). However, compared with Artificial Neural Networks, Decision Trees have not often been applied for land cover classification using Remote Sensing and GIS data. Some applications of Decision Trees for forest mapping in the study area of this thesis are reviewed in section 3.3, and other applications of Decision Trees using remotely sensed data and/or GIS data are reviewed below.

Decision Trees have been used to classify remotely sensed data. For example, a Decision Tree based on a statistics design used the approach of "guided search with forward pruning" to classify remotely sensed data (Swain & Hauska, 1977). The classification was divided into four stages. First, it uses a measure of classifier separability to decide appropriate classes to be classified. Second, it employs a feature selection algorithm to determine a subset of features to be used for branching the Decision Tree. Third, it uses a heuristic search procedure to create the Decision Tree. Finally, it draws the Decision Tree and codes it appropriately for classification. The results of the Decision Tree classification were compared to a single-stage Maximum Likelihood Classifier. The Decision Tree achieved superior performance on both efficiency and accuracy, even though the Decision Tree was only suboptimal. Another Decision Tree algorithm based on the statistics method was reported by Argentiero *et al.* (1982) for the classification of Remote Sensing imagery. The algorithm relies on linear feature extractions and Bayesian look-up table decision rules. Associated error matrices are used to provide an optimal design of Decision Tree at each node. The Decision Tree so created can produce not only a hard classification, but also probability-based soft output. The results showed a classification accuracy of 75% compared to the theoretically optimal 79% with a fully dimensional Bayesian classifier.

However, Decision Trees are more suitable for the classification of multisource data because they are also non-parametric-based models, and they do not need to assume data distribution or need to know data distributions like conventional statistical methods do. Meanwhile, Decision Trees can automatically figure out the internal relationships among data sources, which is an important advantage. Applications which fit into this group include a study to predict Greater Glider Density in Australia using three Decision Trees (CART, ID3, and CN2) (Stockwell *et al.*, 1990), a study using ID3 and CART to determine relationships among lake acidification data (Liepins *et al.*, 1990), a study to predict the Greenness Vegetation Index using CART (Michaelsen *et al.*, 1994), and a study of vegetation classification using CART (Hansen *et al.*, 1996).

Furthermore, besides the Univariate Decision Tree that selects only one feature at each node to split the tree and the Multivariate Decision Tree that uses more than one feature for the splitting rule (Brodley & Utgoff, 1995), the Hybrid Decision Tree, in which different algorithms used in different sub-trees of a large tree, can also be used for classifications. These algorithms could be other Decision Trees, linear discriminant functions, and the *k*-means classifier (Friedl & Brodley, 1997). Friedl and Brodley (1997) concluded that Decision Trees, especially the Hybrid Decision Tree, could produce consistently higher classification accuracies than conventional classifiers.

2.2.3 Problems of Decision Trees

Compared to Artificial Neural Networks, Decision Trees are more stable in performance and less complicated to apply. However, there are still some factors that may affect their classification performance. Pal and Mather (2003) evaluated the effect of the following factors on the classification accuracy of three Decision Trees: training data set size, dimensionality of the data set, attribute selection measures (splitting rule), pruning methods, and boosting techniques. They concluded that the size of the training data set did have positive relationship with the classification accuracy, so did an appropriate pruning method and the boosting techniques. However, they found that the choice of attribute selection measures was not an important factor, and Decision Trees performed relatively poor for high-dimensional data sets.

There are other drawbacks associated with Decision Trees. For example, for a large tree, a feature may be tested several times along a path from root to leaf (e.g., CART), which imposes questions on its comprehensibility. Meanwhile, errors may be accumulated from level to level in a large tree (Wang & Suen, 1987). In particular, data errors associated with the variables which appear in the upper levels of a tree would have significant negative effects on the final results. Decision Trees are sequential in nature, which make them less effective users of training samples than Artificial Neural Networks. Therefore a large number of samples are needed for effectively learning a large tree (Safavian & Landgrebe, 1991; Lees, 1996c). Decision Trees are also very sensitive to sampling errors. In addition, even though optimisation techniques have been suggested for designing Decision Trees (Meisel & Michalopoulos, 1973; Payne & Meisel, 1977; Kurzynski, 1983), most Decision Trees are sub-optimal in nature.

Furthermore, even though Decision Trees are more comprehensible than Artificial Neural Networks, they still can be difficult to process when the intention is to use them in Expert Systems. Algorithms such as PRISM have been suggested to close the gap (Cendrowska, 1987).

2.2.4 Summary of Decision Trees

Decision Trees enjoy several significant advantages over conventional statistical classifiers. Firstly, Decision Trees are strictly non-parametric models. Therefore multisource data with different data distributions could be easily combined in applications. Most Decision Trees have no problem in using all kinds of data types including categorical data, numerical data, and ordinal data. Secondly, Decision Trees are uncertainty tolerant. They are able to deal with noisy and missing data. Moreover, Decision Trees employ automatic feature selection and feature reduction, and they simply ignore irrelevant features. In addition, Decision Trees are more comprehensible to users and experts than conventional statistical classifiers and Artificial Neural Networks. Also, Decision Trees are robust models which do not suffer from the problem of instability associated with many Artificial Neural Networks. There are much fewer factors that can affect the performance of Decision Trees compared with Artificial Neural Networks.

On the other hand, Decision Trees have the disadvantages of a large sample size requirement, error accumulation from level to level, and are sub-optimal in nature when compared with Artificial Neural Networks.

2.3 DEMPSTER-SHAFER'S THEORY OF EVIDENCE

Compared with Artificial Neural Networks and Decision Trees, Dempster-Shafer's theory of evidence (Shafer, 1976) has not received much attention in the classification field. However, Dempster-Shafer's theory is also non-parametric, and it enjoys several advantages for classification problems. Dempster-Shafer's theory is good at handling incomplete (ignorance) and noisy data. It is able to combine new information whenever an item of independent evidence is obtained, and does not require the model to restart as

Artificial Neural Networks and Decision Trees do. Dempster-Shafer's theory is also able to cope with hierarchical classifications.

2.3.1 Principles of Dempster-Shafer's theory

Dempster-Shafer's theory deals with the combination or pooling of independent bodies of evidence (Shafer & Logan, 1987). It is a theory of evidence because it deals with the weights of evidence and with numerical degrees of support based on available evidence (Shafer, 1976). The idea of Dempster-Shafer's theory is to decompose a larger body of evidence into manageable components, assign the mass of support for each of these components, then combine them to produce a total amount of belief (Srinivasan & Richards, 1990). The mass of support could be conditional probability, expert judgement, confidence factor, and even fuzzy membership (Dubois & Prade, 1990). Dempster-Shafer's theory separates itself from Bayes theory by rejecting the rule of additivity, i.e., Bel(A) + Bel(~A) <= 1. Shafer (1976) has argued that it is the rule of additivity that makes Bayes theory incapable of handing *ignorance*. He further pointed out that Dempster's rule of combination is an excellent tool for combining weights of evidence.

In order to better introduce the concepts of Dempster-Shafer's theory, a simple case of land cover classification is described below. Suppose there is a universal set of forest types Θ which is called *frame of discernment*. A subset of Θ is the conifer forest that has two classes, c_1 and c_2. Another subset of Θ is the deciduous forest that has three types of d_1, d_2, and d_3. The third subset of Θ is the broad leaf forest including three types of b_1, b_2, and b_3. Now suppose for a pixel, information from one item of evidence assigns the following *masses of support* or *basic probability assignments* to these forest types; $m(c_1) = 0.2$, $m(d_1) = 0.2$, $m(d_2) = 0.1$, and $m(b_1) = 0.4$. The remaining support was set to Θ, i.e. $m(\Theta) = 0.1$, which is the mass of support of *ignorance*, and it is the uncertainty posed on the assignment. The *plausibility* is measured as; $p(a) = 1 - m(\sim a)$, where ~a is the complement of a. So for this case, $p(c_1) = 0.3$, $p(d_1) = 0.3$, $p(d_2) = 0.2$, and $p(b_1) = 0.5$. The difference between the plausibility and the mass of support is called *evidential interval*, which for this case is 0.1 for all elements. If now, the second item of evidence from an independent source gives the following new masses of support, $m(c_2) = 0.4$, $m(d_3) = 0.1$, $m(b_1) = 0.2$, and $m(b_3) = 0.1$, how can we update the mass of support for each of these forest types, and what is the belief we can assign to each of the three broad forest classes (conifer, deciduous, and broad leaf)? As we have mentioned above, Dempster's rule of combination (orthogonal sum) can be used to combine new obtained independent evidence. The rule can be expressed as algebraic form:

$$m_1 \oplus m_2(Z) = \frac{1}{1-k} \sum_{X \cap Y = Z} m_1(X) m_2(Y) \tag{2.27}$$

where $k = \sum_{X \cap Y = \phi} m_1(X) m_2(Y)$, and m_1 and m_2 are the masses of support from two items of evidence, X and Y are the elements of Θ, $Z = X \cap Y$, and k is the mass that the combination assigns to the null subset $\varnothing$. The rule can also be illustrated as table form. For example, a mass table for the above case study is built as (Table 2.1):

Table 2.1 An example of mass table

mass	c_1	d_1	d_2	b_1	Θ
c_2	$\varnothing$ / 0.08	$\varnothing$ / 0.08	$\varnothing$ / 0.04	$\varnothing$ / 0.16	c_2 / 0.04
d_3	$\varnothing$ / 0.02	$\varnothing$ / 0.02	$\varnothing$ / 0.01	$\varnothing$ / 0.04	d_3 / 0.01
b_1	$\varnothing$ / 0.04	$\varnothing$ / 0.04	$\varnothing$ / 0.02	b_1 / 0.08	b_1 / 0.02
b_3	$\varnothing$ / 0.02	$\varnothing$ / 0.02	$\varnothing$ / 0.01	$\varnothing$ / 0.04	b_3 / 0.01
Θ	c_1 / 0.04	d_1 / 0.04	d_2 / 0.02	b_1 / 0.08	Θ / 0.02

The updated masses of support after combining the two items of evidence using equation 2.27 are; $m(c_1) = 0.04/(1-0.64) = 0.11$, $m(c_2) = 0.11$, $m(d_1) = 0.11$, $m(d_2) = 0.056$, $m(d_3) = 0.0278$, $m(b_1) = 0.5$, $m(b_3) = 0.0278$, and $m(\Theta) = 0.056$. Meanwhile, Bel(A) is a measure of degree of belief assigned to the subset A, and is the sum of the masses of support assigned to all elements of A. For example,

Bel(conifer) = $m(c_1) + m(c_2) = 0.22$,

Bel(deciduous) = $m(d_1) + m(d_2) + m(d_3) = 0.1938$, and

Bel(broad leaf) = $m(b_1) + m(b_2) + m(b_3) = 0.5278$.

Hard decisions can be made on any of the following four rules (Lee *et al.*, 1987):

1. A maximum support rule, which chooses the proposition with the highest mass of support;
2. A maximum plausibility rule, which chooses the proposition with the highest plausibility;

3. An absolute rule, which chooses the proposition whose mass of support is greater than the plausibilities of all other propositions; and
4. A maximum support and plausibility rule, which chooses the proposition with both the highest support and plausibility.

2.3.2 Applications of Dempster-Shafer's theory

There are many fewer applications of Dempster-Shafer's theory in the classification field to date. The reasons may be due to little awareness of the theory by many researchers in the classification field, the complexity of the algorithm, and the lack of any commercial package available for automatic implementation. Dempster-Shafer's theory has many advantages in dealing with the classification problem over conventional statistical classifiers.

Dempster-Shafer's theory is good at handling situations when there is conflicting evidence arising from different data sources, when there is ignorance in data sources, when multisource data is involved, and when data sources are not based on a common frame of discernment (e.g., hierarchical classification). These advantages have been demonstrated by studies of Bogler (1987), Lee *et al.* (1987), Peddle & Franklin (1992), Peddle & Franklin (1993), Peddle (1995), and Gong (1996), when compared with conventional statistical methods. These studies have shown that ignorance could be suspended and the decision deferred until a later stage, and that Dempster-Shafer's theory was better in dealing with uncertainty.

Dempster-Shafer's theory is also good at dealing with incomplete and noisy data. While incomplete and noisy data impose great difficulties on the implementation of formal probability assignments, Dempster-Shafer's theory can use subjective beliefs assigned by domain experts. One such application is reported by Moon (1990) for identifying metal deposits. In the study, he emphasized that "ignorance" is different from "disbelief" in that ignorance shows the degree to which the proposition is uncertain, while disbelief shows the degree to which the proposition cannot be believed.

The hierarchical classification problem is another application area of Dempster-Shafer's theory. For example, the studies of Wilkinson and Megier (1990) and Kontoes *et al.*

(1993) have demonstrated the usefulness of Gordon-Shortliffe's approximation of Dempster-Shafer's algorithm (Gordon & Shortliffe, 1985) in dealing with the hierarchical classification problem.

2.3.3 Problems of Dempster-Shafer's theory

One significant drawback of Dempster-Shafer's theory is the exponential increase of computational complexity when items of evidence are linearly increased. This is especially the case when hierarchical evidences are involved. Several modifications of the initial theory have been recommended to ease the problem. For example, Barnett (1981) has suggested that if we restrict the hypotheses of interest to the mutually exclusive singletons and their negations, the computation time only increases linearly with the increase of items of evidence. Gordon-Shortliffe's approach (Gordon & Shortliffe, 1985), meanwhile, requires an assumption that the hypotheses space can be reduced to a strict hierarchy, and an approximation to assign disconfirming beliefs only to hypotheses with 'meaning' in the domain. It does actually achieve a computationally tractable execution time in managing hierarchical evidence. The problem is that the belief interval (the evidential interval) is lost in the scheme. Shafer and Logan (1987) later argued that Gordon-Shortliffe's approximation is weak in certain circumstances. So, they suggested a new algorithm for the exact implementation of Dempster's rule so that Gordon-Shortliffe's approximation can be removed, which requires less computation time than Gordon-Shortliffe's approach. Vookbraak (1989) also suggested another computationally efficient Bayesian approximation of Dempster-Shafer's theory.

Another drawback of Dempster-Shafer's theory is that unreasonable results may be obtained in the combination of conflicting items of evidence. This problem was first discovered by Zadeh (1984), when a combination assigned 100% certainty to a minority opinion. Other problems also associated with the combination of conflicting items of evidence, include the loss of evidential interval, and the gains of a disproportionate share of belief with elements that have larger cardinality (Murphy, 2000). These problems are due to the normalisation process of Dempster's rule. Several options have been suggested to deal with the problems:

1. Allowing mass of support in the null set instead of updating the mass from normalisation;

2. Assigning the mass of null set to Θ; and
3. Averaging the masses assigned to a subset Z to determine its belief function.

Murphy's study (2000) has shown that each of the three options could ease these problems, but the average method was the best because it preserved the records of uncertainty and relative beliefs.

2.3.4 Summary of Dempster-Shafer's theory

Although some people argue that Dempster-Shafer's theory is only a special case of Bayes theory (i.e., Kyburg, 1987; Lindley, 1987), and many critiques exist for the theory (e.g., Peal, 1990; Vorrbraak, 1991), Dempster-Shafer's theory does have several significant advantages over probability theory. Advantages of the theory pointed out by Srinivasan and Richards (1990) include distinguishing between lack of belief (ignorance) and disbelief, interval of belief, and suspension of judgement to a later stage. Other advantages over Bayes theory are its ability to cope with different data types, no assumption of data distribution, and its capability for handling uncertain data. Some attractive features of Dempster's rule of combination are; concordant items of evidence reinforce each other; conflicting items of evidence erode each other; and a chain of reasoning is weaker than its weakest link (Shafer, 1990; Murphy, 2000).

Two major problems highlighted above are associated with computational complexity and difficulties when combining conflicting items of evidence. However, when the aim is to discriminate mutually exclusive classes the problem of computational complexity can be largely eased.

One assumption made by Dempster-Shafer's theory is the independence of items of evidence, which is not usually satisfied for a real problem and therefore brings errors into the classification process. But it may serve as an advantage sometimes (e.g., Lee *et al.*, 1987). The assumption of independence also avoids biasing the modelling process to any input variable that may have high-level data error.

2.4 FUZZY SET THEORY

Is the world exact? It is definitely not! Our world is full of uncertainties. Traditionally, Bayes theory or probability theory is used to handle the uncertainties. But it only deals with the randomness of an event, and completely ignores other aspects an event may have, such as ambiguity and vagueness. The inability of Bayes theory to handle these kinds of uncertainties has worried a large number of researchers until the development of fuzzy set theory by Zadeh (1965). Fuzzy set theory is designed to deal with the uncertainties associated with inexact event and phenomena, and has an excellent theoretical foundation for handling issues such as fuzzy linguist variables, fuzzy models, and fuzzy logic.

2.4.1 Principles of fuzzy set theory

Zadeh (1993) defined the theory of fuzzy sets as "a body of concepts and techniques for dealing in a systematic way with a type of imprecision which arises when the boundaries of a class of objects are not sharply defined." To recognise the usefulness of fuzzy set theory, it is necessary to separate fuzziness from randomness (Zadeh, 1993). As pointed out by Kosko (1990), fuzziness measures the degree to which an event occurs, while randomness concerns whether or not it occurs and if so, to what degree. For example, what chance tomorrow will be raining is randomness, but to what degree the rain is heavy is fuzziness. Randomness occurs when there is not enough information to resolve the uncertainty, and it gradually disappears with the increasing information available (Kosko, 1990). However, fuzziness resides deep in the nature of an event or in the definition of an inexact phenomenon, it has nothing to do with the information or knowledge we have on the event or the phenomenon. Fuzzy set theory or possibility theory differs from probability theory by not obeying the law of noncontradiction and the law of excluded middle. For example, in fuzzy set theory, it is normal to have $A \cap \sim A \neq \emptyset$, and $A \cup \sim A \neq U$, where U is a *universal set.* But there does exist a *possibility/probability consistency principle* (Zadeh, 1978), which indicates that an impossible event is always improbable, but an improbable event does not necessarily mean impossible.

To appreciate fuzzy set theory, basic notions need to be introduced (Zadeh, 1993). *Membership function* is a function that describes the degree of belonging of a fuzzy subset A to a universal set U, $\mu_A : U \rightarrow [0,1]$. Frequently used membership functions include S-function, $\prod$-function, and Z-function, as defined and plotted below:

1. S-function (Figure 2.2a)

$$S(u;a,b,c) = \begin{cases} 0 & u \le a \\ 2 \times (\frac{u - a}{c - a})^2 & a \le u \le b \\ 1 - 2 \times (\frac{u - c}{c - a})^2 & b \le u \le c \\ 1 & u \ge c \end{cases} \tag{2.28}$$

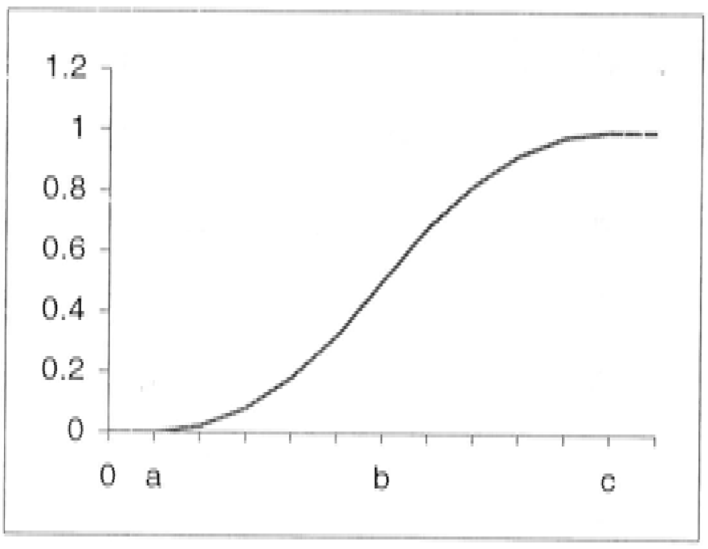
where $b = \frac{a+c}{2}$.

2. $\prod$-function (Figure 2.2b)

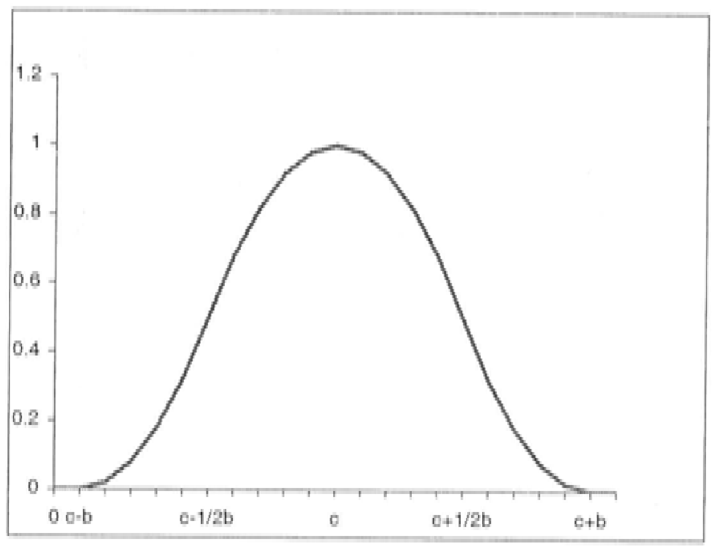
$$\prod(u;b,c) = \begin{cases} S(u;c-b,c-\frac{1}{2}b,c) & u \le c \\ 1 - S(u;c,c+\frac{1}{2}b,b+c) & u \ge c \end{cases} \tag{2.29}$$

3. Z-function (Figure 2.2c)

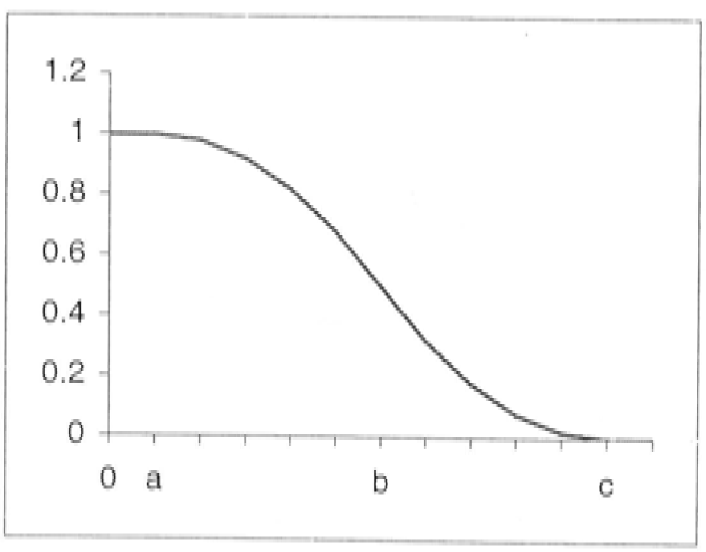
$$Z(u;\ a,\ b,\ c) = 1 - S(u;\ a,\ b,\ c) \tag{2.30}$$



(a)



(b)



(c)

Figure 2.2 Fuzzy membership functions, (a) S-function, (b) $\prod$-function, (c) Z-function

The *grade of membership* is represented as $\mu_A(u)$. The *support* of A is the set of points in U at which $\mu_A(u)$ is positive. The *height* of A is the supremum of $\mu_A(u)$. The *crossover point* of A is the point that has grade of membership of 0.5. And A is *normal* if its height is unity, and *subnormal* in other cases.

A finite fuzzy subset can be represented as:

$$A = \sum_i \mu_i / u_i \tag{2.31}$$

where $1 \leq i \leq n$

And for arbitrary fuzzy subset, it is represented as:

$$A = \int \mu_A(u) / u \tag{2.32}$$

A *Type 1* fuzzy subset is that its membership function is a mapping from U to [0,1], while *Type 2* is a mapping from U to the fuzzy subsets of *Type 1*. For example, $\mu_{small\ integer}(u) = 1/1 + 0.8/2 + 0.6/3 + 0.4/4 + 0.1/5$ is a *Type 1* fuzzy subset, while $\mu_{small\ integer}(5) = low$ is a Type 2 fuzzy subset, because *low* is itself a fuzzy subset. It is also defined that *α-cut* of a fuzzy subset is all elements of U whose grade of membership in A is greater than or equal to α.

The basic operations of fuzzy set theory are very different from those of probability theory. They are reviewed below (Zadeh, 1993).

1. The *complement* of A is:

$$\sim A = \int (1 - \mu_A(u)) / u \tag{2.33}$$

2. The *union* of fuzzy sets A and B is:

$$A \bigcup B = \int (\mu_A(u) \vee \mu_B(u)) / u \tag{2.34}$$

where $\vee$ is maximum operation.

3. The *intersection* of fuzzy sets A and B is:

$$A \bigcap B = \int (\mu_A(u) \wedge \mu_B(u)) / u \tag{2.35}$$

where $\wedge$ is minimum operation.

4. The *product* of A and B is:

$$A \times B = \int (\mu_A(u) \times \mu_B(u)) / u \tag{2.36}$$

$A^2 = \int (\mu_A(u))^2 / u$ is called *concentration*, and

$A^{0.5} = \int (\mu_A(u))^{0.5} / u$ is called *dilation*.

5. The *bounded-sum* of A and B is:

$$A \oplus B = \int 1 \wedge (\mu_A(u) + \mu_B(u)) / u \tag{2.37}$$

6. The *bounded-difference* of A and B is:

$$A \ominus B = \int 0 \vee (\mu_A(u) - \mu_B(u)) / u \tag{2.38}$$

7. The *left-square* of A is:

$$^2A = \int \mu_A(u) / u^2 \tag{2.39}$$

8. If $A_1, A_2, A_3, \cdots, A_n$ are fuzzy subsets of $U_1, U_2, U_3, \cdots, U_n$, the *cartesian product* of $A_1, A_2, A_3, \cdots, A_n$ is:

$$M_{A_1 \times A_2 \times \cdots \times A_n}(U_1, U_2, \cdots, U_n) = \mu_{A_1}(u_1) \wedge \mu_{A_2}(u_2) \wedge \cdots \wedge \mu_{A_n}(u_n) \tag{2.40}$$

9. *Fuzzy relation* of $U_1, \cdots, U_n$ is:

$$R = \int_{U_1 \times \cdots \times U_n} \mu_R(u_1, \cdots, u_n) / (u_1, \cdots, u_n) \tag{2.41}$$

10. The *extension principle* is expressed as:

$$f(A) = \int \mu_A(u) / f(u) \tag{2.42}$$

where $f(A)$ is a mapping from U to the other universal set of V. For example, the left-square (equation 2.39) is a fuzzy subset of this kind.

11. *Convex combination* of $A_1, \cdots, A_n$ is:

$$\mu_A = w_1 \mu_{A_1} + w_2 \mu_{A_2} + \cdots + w_n \mu_{A_n} \tag{2.43}$$

where $\sum_i w_i = 1$, and $w_1, \cdots, w_n$ are nonnegative weights.

Fuzzy subsets can be modified by combining modifiers, qualifiers, and quantifiers. The *modifiers* such as <u>not</u>, <u>very</u>, <u>somewhat</u>, <u>more or less</u>, and <u>slightly</u> are used to modify the attribute of a fuzzy variable. For example, "John is young" could be modified as "John is <u>very</u> young". The general rule is from "X is F" to "X is <u>m</u>F", where '<u>m</u>' is a modifier. Among these often used modifiers, <u>not</u> is the complement, <u>very</u> is the concentrator, and <u>more or less</u> is the dilator. <u>Somewhat</u> is often defined as $F^{0.333}$, and <u>extremely</u> is often defined as F^3. The *qualifiers* are truth-values, likelihood-values, and possibility-values. For example, "John is young" could be translated to "John is young is <u>very true</u>" or "John is young is <u>quite likely</u>" or "John is young is <u>highly possible</u>". Meanwhile, the *quantifiers* such as <u>several</u>, <u>a few</u>, <u>at least 5</u>, <u>about 5</u> are used to quantitatively measure the fuzzy propositions. For example, in the sentence of "there are <u>several</u> people in the room", '<u>several</u>' is a quantifier. It may be defined as: $\mu_{several}$ =

0.2/1 + 0.4/2 + 0.6/3 + 0.8/4 + 1/5 + 1/6 + 0.6/7 + 0.3/8. The membership functions of these modifiers, qualifiers, and quantifiers must be defined by users in a case-by-case basis.

Besides Zadeh's view of *sets-as-functions*, Kosko (1990) pointed out that fuzzy set theory could be seen from the geometric view as *sets-as-points*. The view sees a fuzzy set as a unit hypercube, and a fuzzy subset of the fuzzy set is a point in the hypercube, while vertexes are non-fuzzy subsets of the fuzzy set. The midpoint of the hypercube represents the maximally fuzzy, and it is a point forbidden by classical logic and set theory. He also used fuzzy Hamming distance to measure "how big is a fuzzy set", and used fuzzy entropy theorem to measure "how fuzzy is a fuzzy set". It was concluded that the geometric view of fuzzy sets was very useful in understanding fuzziness, defining fuzzy concepts, and proving fuzzy theorems.

2.4.2 Applications of fuzzy set theory

Since its establishment in 1965, applications of fuzzy set theory and its extension, fuzzy logic, have blossomed in almost all fields of science, technology, management, and decision-making. Commercial exploitation of fuzzy industrial controllers has been most successful in the past two decades. Maiers and Sherif (1985) have reviewed a larger number of applications of fuzzy set theory in all of these fields. This section, however, focuses on only applications of fuzzy set theory in the geoscience field, especially those applications of fuzzy linguistic variables and applications of fuzzy classifications.

2.4.2.1 Applications of fuzzy set theory in the geoscience field

The geographical world is full of fuzziness. Fisher (2000) has used the *paradox of Sorites* to test whether a geographical concept is vague or not. He found that not only geographical relations such as proximity and direction are vague, but also geographical objects such as urban and rural and geographical processes such as seasonal change of vegetation are vague. Therefore, he concluded that vagueness is endemic in geographical thinking and in geographical information. Then, he went on to suggest using fuzzy set theory to handle all these vague concepts.

Concepts of fuzzy linguistic variables have been used to define fuzzy spatial relations of distance, direction and neighbourhood. For example, fuzzy distance measurements such as short are an important spatial concept in our daily life, and have been defined by quite a few researchers (e.g. Leung, 1982; Altman, 1994; Albercht & Guesgen, 1998; Robinson, 2000). Meanwhile, Leung (1982) and Altman (1994) have also defined the concepts of fuzzy directions such as a bit north. Moreover, definitions of fuzzy neighbourhood concepts such as close to can be found in Leung (1982), Albrecht and Guesgen (1998), Guesgen and Albrecht (2000), and Cobb *et al.* (2000).

The boundary between regions is one of the most obvious fuzzy spatial objects. Natural phenomena such as vegetation types usually impose gradual changes across a continuum; man-made definitions of sharp boundaries between them inevitably cause uncertainty. To deal with the problem, Leung (1987) defined the core of a region which is the point or area in the region whose characteristics are most compatible to the region, the edge of the region which are adjacent points in the region whose characteristics completely disappear, and the boundary of the region which are the points whose characteristics are more or less compatible to the region, or area between the core and the edge of the region. He defined the boundary between two regions as a zone instead of as a sharp line, which includes all points whose characteristics are more or less compatible to both of the regions. Meanwhile, Wang and Hall (1996) described the fuzzy representation of geographical boundaries based on the degree of sharpness. The degree of sharpness is defined as a grade of membership which represents change sharpness at a boundary of interest, and is defined in terms of the first order derivative. They concluded that the fuzzy boundaries so defined could describe not only the location but also the rate of change of environmental phenomena at or about the boundary. Other applications of fuzzy objects which need to be mentioned are a fuzzy object model developed by the Fuzzy Object Data Management Group through the integration of two techniques; fuzzy set theory and object data modelling (Cross & Firat, 2000) and three fuzzy object models (Fuzzy-Fuzzy object, fuzzy-Crisp object, and Crisp-Fuzzy object) proposed by Cheng *et al.* (2001) to represent objects with fuzzy spatial extents.

Fuzzy geographical processes can be modelled by fuzzy set theory. For example, Dragicevic and Marceau (2000a, 2000b) described an application of fuzzy set theory in modelling a very dynamic rural-urban environment of Montreal Metropolitan area in

Quebec, Canada, from 1956 to 1986 with a temporal resolution of 10 years. The datasets are four geo-registered land-use maps at snapshots of 1956, 1966, 1976, and 1986. Fuzzy set theory was used to perform temporal interpolation between these snapshots. The results indicated that fuzzy set theory could generate realistic temporal interpolation of urban expansion process based on different scenarios, and it was able to model the time change of the dynamic phenomenon conveniently.

Fuzzy set theory is also a very useful tool for building fuzzy queries of GIS. For example, a fuzzy query language FQUEL has been designed to incorporate fuzzy statements, fuzzy relational operators, fuzzy connectors and fuzzy modifiers in a Fuzzy Relational Soil Information System (FRSIS) (Kollias & Voliotis, 1991). Meanwhile, fuzzy query to integrate more natural language expressions into GIS user interface was suggested by Wang (1994). The approach enables the use of modification-type (modifiers), composition-type, and quantification-type (quantifiers) fuzzy formulae for the data retrieval. A defuzzification process to obtain nonfuzzy outputs from using fuzzy queries was also described. The approach was later enhanced by combining with fuzzy grammar theory (Wang, 2000). The natural language interface so created can process queries of simple English sentences that are grammatically correct. Stefanakis *et al.* (1999) conducted a study on why and how to incorporate fuzzy set theory into GIS. Methods were suggested to extend the standard GIS spatial operations, spatial measures, and spatial queries to support fuzzy representations and fuzzy reasoning.

During the past two decades, there have been increasing number of applications of fuzzy set theory to real world geographical problems. For example, there are applications in environmental and GIS modelling (e.g. Zhu *et al.* 1996; Urbanski, 1999; Mackay & Robinson, 2000), applications in land suitability analysis (e.g. Hall *et al.*, 1992; Wang *et al.*, 1990; Banai, 1993; Jiang & Eastman, 2000), applications on slope stability prediction (e.g., Davis & Keller, 1997), and applications on the accuracy assessment of thematic maps (e.g., Gopal & Woodcock, 1994; Woodcook & Gopal, 2000; Townsend, 2000; Power *et al.*, 2001; Laba *et al.*, 2002; Hagen, 2003).

2.4.2.2 Applications of fuzzy classification

Fuzzy classification is an active area of Remote Sensing and GIS application. The essential idea of fuzzy classification is to derive fuzzy membership functions that can produce soft classification instead of hard classification. Thus, fuzzy classification is the most suitable approach in dealing with the mixed pixel problem. Two methods are believed to be the most appropriate in deriving fuzzy membership functions. Fuzzy *c*-means or fuzzy *k*-means is based on a clustering principle, while the semantic import model is based on the expert or empirical model (Burrough, 1989; Fisher, 2000).

The fuzzy *c*-means algorithm was first developed by Bezdek *et al.* (1984). This program generates fuzzy partitions for corroborating known substructures or suggesting substructure in unexplored data. The algorithm uses a generalized least-square function to aggregate subsets. It is basically unsupervised, although a supervised version can be implemented as well. Applications of the fuzzy *c*-means classification can be found in Trivedi & Bezdek (1986), Cannon *et al.* (1989a, 1989b), Fisher and Pathirana (1990), McBratney & Gruijter (1992), Foody (1992), Foody (1994), Gruijter *et al.* (1997), Franssen *et al.* (1997), Irvin *et al.* (1997), Burrough *et al.* (1997), Burrough *et al.* (2000), and Zhang & Stuart (2001). Fuzzy classifications based on the semantic import model can be found in Burrough *et al.* (1992), and Doberman and Oberthur (1997). Artificial Neural Networks could be used to produce fuzzy or soft classification (Foody, 1996; Warner & Shank, 1997; Foody, 1999). In addition, besides the fuzzy *c*-means algorithm, Statistical methods (i.e., the Maximum Likelihood Classifier) have been used to derive fuzzy membership functions for the purpose of fuzzy or soft classifications (e.g., Kent & Mardia, 1988; Wang, 1990a; Wang, 1990b; Foody *et al.*, 1992; Canters, 1997; Shackelford and Davis, 2003). However, some of these applications are not strictly fuzzy classifications but soft classifications, as they use probability theory instead of fuzzy set theory for the applications.

2.4.3 Problems of fuzzy set theory

There are some disadvantages associated with fuzzy set theory. First, it is not necessarily the only or the best way for handling uncertainties. Fuzzy set theory is not appropriate for dealing with randomness and ignorance. Second, it is often difficult to derive sensible and consistent fuzzy membership functions, even though some experimental and optimise acquisition methods have been suggested (Turksen, 1991; Bagis, 2003). This is the most difficult problem of fuzzy set theory. The process of deriving fuzzy membership functions often involves subjective. Moreover, it is sometimes inconvenient to implement fuzzy set theory in real world problems, as it often involves troublesome fuzzification and defuzzification processes.

2.4.4 Summary of fuzzy set theory

Fuzzy set theory is a break from the classical Bayes theory, which marks the beginning of new views on engineering control, the decision-making process, and scientific research. Fuzzy set theory is an appropriate tool for representing fuzzy information, describing fuzzy events or objects, modelling fuzzy relations, simulating fuzzy processes, and implementing fuzzy reasoning. Fuzzy set theory is versatile, flexible, soft, and intelligent. In addition, even though it is difficult to derive sensible and consistent fuzzy membership functions, expert knowledge can help in most cases.

2.5 EXPERT SYSTEMS

What are Expert Systems? A useful and simple definition may be "programs that simulate the decision making process of domain experts, and actually achieve decisions of expert level in any circumstance". But a broader definition should include all programs that manipulate knowledge rather than numbers (Leary, 1988). For over 20 years, Expert Systems have been a focus of AI research, with a huge number of applications across nearly all scientific fields including engineering, chemistry, medicine, geology, computing, and planning (Giarratano & Riley, 1998). Expert Systems have been commercialised since the 1980s, and some of them have been playing important roles on the daily running of several organisations, such as the XCON Expert System for computer configuration (McDermott & Bachant, 1984) (see Crevier, 1993; Chapter 3 for details).

2.5.1 Principles of Expert Systems

A typical Expert System is consist of the following components (Figure 2.3):

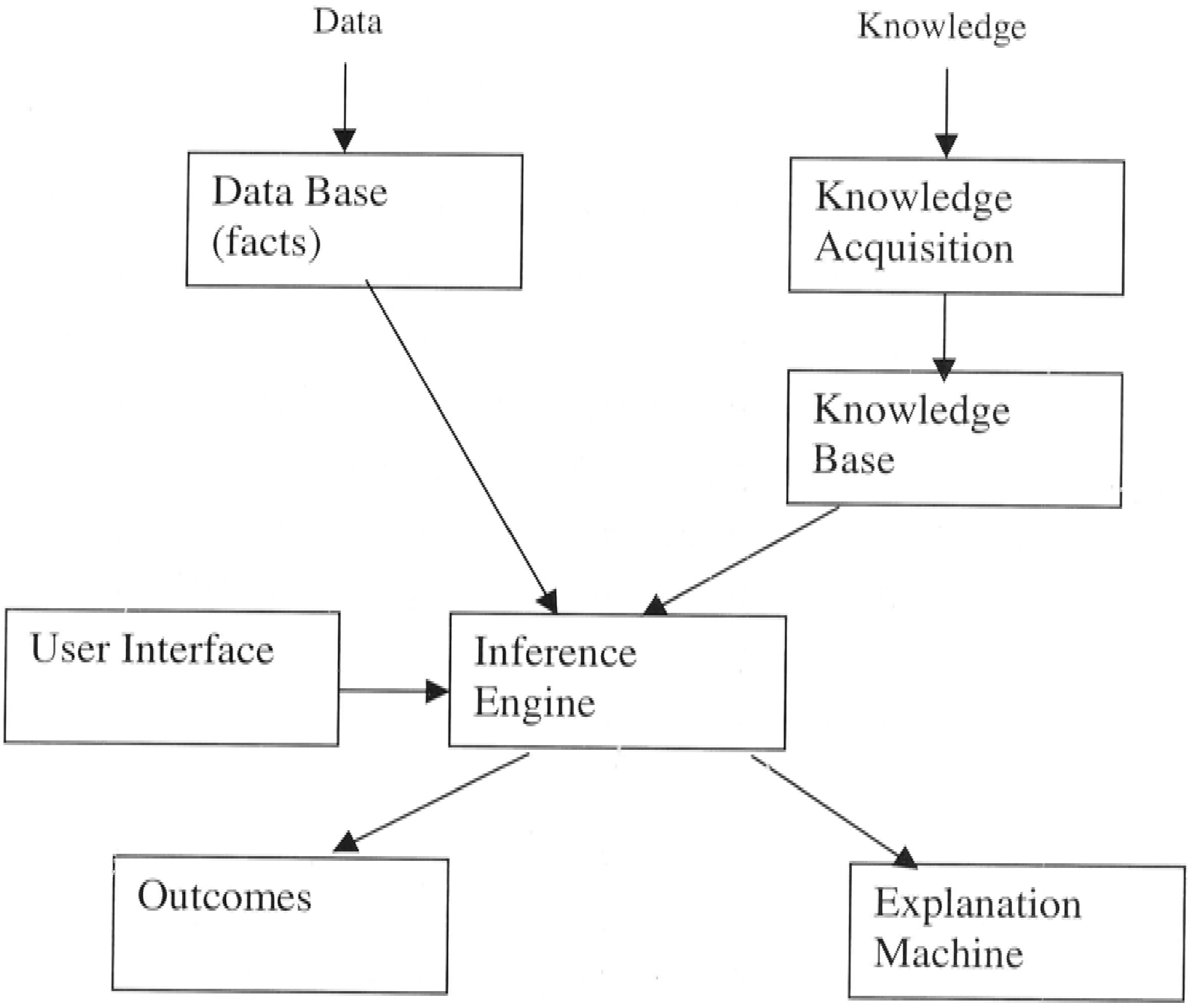


Figure 2.3 Typical components of an Expert System

Among these components the data base, the knowledge base, the inference engine, and the explanation machine are the essential components. The data base stores the data of input variables, and provides facts for the inference engine. The knowledge base, on the other hand, stores expert knowledge in some appropriate styles such as rules, frames, and so on. It provides to the inference engine the domain knowledge derived from experts. The inference engine matches facts with one piece or a group of pieces of knowledge, and induces conclusion. The explanation machine however, explains the reasoning behind the inference process and how the conclusions are obtained.

Expert knowledge in Expert Systems can be represented as production rules, semantic networks, frames, schema, or scripts (Partridge, 1996; Giarratano & Riley, 1998). All of them have advantages and disadvantages. The best choice must be assessed case by case. For example, Ramsey *et al.* (1986) made a detailed comparison of three such methods based on probability theory, the production rule method, and the frame approach. They found that none of them was obviously the best method. A major advantage of the probability-based method is that it has been repeatedly used with proven success. In addition, it is very easy to organize and implement the knowledge base if all needed probabilities are available. But the problem is that it is not uncommon that these probabilities are not all readily available. Meanwhile, unrealistic assumptions on data distribution are involved in applying the approach. The rule-based method has also many proven successes, and it provides the ability to chain associative information to make deductions which can be very useful in a decision-making process. Not only quantitative data but also qualitative knowledge can be handled easily in rule-based Expert Systems. One of the difficulties associated with the rule-based method is that experts are often not comfortable when asked to describe knowledge as rules. One advantage of the frame-based method is that for many applications, frames are easy and natural to write. Meanwhile, context-dependent information can be handled smoothly, as they are all placed in one frame. One major drawback of the frame-based method is, however, that it is the most experimental of the three methods. Therefore more research on several technical issues is needed. Consequently, Ramsey *et al.* (1986) argued that the criteria to choose a most suitable knowledge representation method depends on three main factors; the pre-existing format of the application knowledge, the type of classification desired, and the amount of context-dependence inherently present in a problem.

Knowledge acquisition is regarded as the most difficult part in building an Expert System, and it is usually stated as the "bottleneck" problem (e.g., Mingers, 1986; Giarratano & Riley, 1998). Traditionally, knowledge acquisition is done by knowledge engineers interviewing domain experts. This process requires close cooperation between the domain experts and the knowledge engineers, in which the knowledge engineers obtain relevant domain knowledge from the experts, and transfer them into computer codes of the Expert System. This is a very time and resource intensive job, and errors are easily made in the process. Of course, knowledge acquisition can also be done by using task analysis methods and special task methods (Zhu *et al.*, 1996). But the knowledge obtained by such methods is shallow knowledge, because it is based on heuristic knowledge instead of causal knowledge (Giarratano & Riley, 1998). The other group of knowledge acquisition methods is so called machine learning or self-learning methods, which learn knowledge from examples (Rendell *et al.*, 1989). Decision Trees and Artificial Neural Networks are two of such models. For example, a Decision Tree of

C4.5 and modified Decision Trees of ID3 have been used for knowledge acquisition (Mingers, 1986; Mingers, 1987a, 1987b; Cendrowska, 1987; Huang & Jensen, 1997). Meanwhile, Castro *et al.* (2001) also suggested a fuzzy machine learning technique in the knowledge acquisition process. The Analytical Hierarchy Process has also been employed to capture expert knowledge for a land capability assessment (Itami *et al.*, 2000). In addition, Wang and Mendel (1992) have developed a method to generate fuzzy rules by learning directly from examples.

The inference engine is used to obtain relations between facts and conclusions, not through algebraic equations or theorems, but through a symbolic reasoning process. Possible reasoning methods for Expert Systems are; deductions, induction, intuition, heuristics, generate and test, abduction, default inference, autoepistemic, non-monotonic inference, and analogy (Giarratano & Riley, 1998). The traditional deductive logic, Boolean logic, is the most classic one that enjoys a thousand years of dominance in the western world. It normally includes a premise (antecedent) and a conclusion (consequence). If a fact matches the premise, then the conclusion is deduced. Nowadays, Boolean logic is losing ground, because partial matching is not allowed in the reasoning process. Therefore, methods for approximate reasoning or inexact reasoning have emerged quickly in the past few decades. Some introduction of approximate reasoning is presented in section 2.5.3.

The explanation machine is used to answer "why", "how" and "what if" questions (Davis *et al.*, 1977; Leary, 1988). Thus Expert Systems can respond to user queries such as "why you need this information?", "how is the conclusion obtained", and "what if I change the fact to…?". The explanation component is necessary for an Expert System to become user-friendly, and sensible. As Strat and Lowrance (1989) have stated, "one of the most highly touted virtues of knowledge-based expert systems is their ability to construct explanations for their lines of reasoning".

2.5.2 Rule-based Expert Systems

Rule-based Expert Systems are the most common form of Expert Systems. This is used in this study, so some introduction to rule-based Expert Systems are useful.

Production rule (system) can be represented as: *If* ··· *Then* style. For example, "*If* it is raining outside *Then* carry an umbrella" is a production rule. In the rule, "it is raining outside" is called antecedent or conditional part or pattern part or Left-Hand-Side, while "carry an umbrella" is called consequence or conclusion or Right-Hand-Side. Domain knowledge can be expressed as production rules. Upon receiving a fact from the database, the inference engine compares the fact with the antecedents of all production rules, places the rules which have been matched into an agenda, ranks these rules with some priority criteria, then executes the rules in the priority order. For example, if a fact "it is raining outside" is received, then the action of "carrying an umbrella" must be taken. The inference engine also has the responsibility to resolve any potential contradiction of the rules and to reason properly in an uncertain environment.

Two chain-reasoning approaches are often employed for the inference engine; they are forward chaining and backward chaining. Forward chaining is data driven, which reasons from facts to conclusions. It is a bottom-up structure and is used in the LISP Expert System language (McCarthy et al, 1965). Backward chaining is goal driven, which reasons from hypothesis to facts that support the hypothesis. It is also called a top-down structure and is used in the PROLOG Expert System language (Roussel, 1975). Forward chaining is said to be suitable for fields of planning, monitor, and control, while backward chaining is suitable for diagnosis. Explanation is easier when using backward chaining than when using forward chaining (Giarratano & Riley, 1998).

How to handle uncertainty appropriately is a big issue in rule-based Expert Systems, because uncertainties are always involved in the database, in the knowledge base, and in the reasoning process. Many approaches have been recommended and used for the problem (e.g., those reviewed by Prade, 1985). For example, the classical rule-based Expert System of MYCIN has employed a "certainty factor" approach (Davis *et al.* 1977; Shortliffe & Buchanan, 1985), which attaches certainty factors to the Right-Hand-Side of rules. Another classical example is PROSPECTOR Expert System that used Bayes theory to handle uncertainty (Hart *et al.*, 1978; Duda *et al.*, 1979). Other options are Dempster-Shafer's theory (Shafer, 1987), which can handle uncertainty of ignorance (see section 2.3 for details), fuzzy set theory or possibility theory (Zadeh, 1983; Farreny & Prade, 1986), which can handle uncertainty of fuzziness (see section 2.4 for details), and the theory of coherent lower provisions (Walley, 1996). All of these approaches have advantages and disadvantages (e.g., Lee *et al.*, 1987; Giarratano & Riley, 1998), and should be assessed case by case. Even though some researchers have argued that the only satisfactory description of uncertainty for Expert Systems is probability (i.e., Lindley, 1987), different methods can be successfully applied to similar applications. For example, for medical applications, there are the certainty factor approach of MYCIN (Shortliffe & Buchanan, 1985), the probability-based approach (Spiegelhalter, 1987), and the fuzzy-based approach (Lesmo *et al.*, 1993).

2.5.3 Principles of Fuzzy (Rule-Based) Expert Systems

Fuzzy Expert Systems are Expert Systems that use fuzzy logic instead of Boolean logic. The concept was first declared by Zadeh (1983). The major purpose of Fuzzy Expert Systems is to effectively manage uncertainty for Expert Systems. If an Expert System has any of the following three conditions, it should be regarded as a Fuzzy Expert System (Zadeh, 1983):

1. The fuzziness of antecedents and/or consequence exist in rules, for example, "*If* X is small *Then* Y is large with CF=0.8" is a fuzzy rule, because "small" and "large" are two fuzzy terms, and "CF=0.8" means that the certainty factor attached to the rule is 0.8;
2. Partial matching between a fact and the antecedent of a rule. For example, provided a fact of "X is very small", it only partially matches the antecedent of the above rule; and
3. The presence of fuzzy quantifiers in the antecedents and/or the consequence of a rule, for example a rule such as "*If* X is small *Then* Y is large is likely", where "likely" is a fuzzy quantifier.

Therefore, both a fuzzy database and a fuzzy knowledge base may exist in a Fuzzy Expert System.

Under the circumstances of a fuzzy database, fuzzy knowledge and partial matching, approximate reasoning instead of exact reasoning becomes the only appropriate inference method for Fuzzy Expert Systems (Zadeh, 1975; Yager, 1980; Yager, 1984; Prade, 1985). One approach using such approximate reasoning is based on fuzzy logic, or the rule of generalized *modus ponents* (Zadeh, 1983). The inference engine of a Fuzzy (rule-based) Expert System is responsible for handling partial matching, deducing conclusions and resolving any contradiction of rules. To deal with partial matching, several methods can be used such as the similarity and proximity measures suggested by Zemankova (1993), and the weighted matching index introduced by Dubois *et al.* (1993) if the antecedents have relative importance among them. Meanwhile, for the deduced conclusion, the MIN and the PRODUCT operations are often used to calculate its certainty factor. The MIN operation takes the minimum certainty factor from partial matchings as the certainty factor of the deduced conclusion. While the PRODUCT operation takes the product of all certainty factors from partial matchings as the certainty factor of the deduced conclusion. Whalen and Schott (1983) have reviewed eight such implication operators. Moreover, when several rules have same antecedents but with different conclusions or conclusions with different certainty factors, a potential contradiction of the rules may happen. Usually some kind of evidence combination approach must be applied to resolve the contradiction and to obtain a single conclusion. Approaches based on probability theory, fuzzy set theory and Dempster-Shafer's theory are possible options. For example, an integrated approach based on possibility theory was reported by Lesmo *et al.* (1985).

Bonde (2000) divided the fuzzy inference process (approximate reasoning) into 4 subprocesses; fuzzification, inference, composition and defuzzification, while the defuzzification subprocess is optional. The fuzzification subprocess uses fuzzy membership functions to calculate the degree of truth for antecedents. The inference subprocess applies the MIN or the PRODUCT operations to determine the truth-value of consequences. In the composition subprocess, all of the rules with the same conclusion but with different degrees of truth-value are combined to form a single rule. The MAX and the SUM operations are most useful methods in this subprocess. In the defuzzification subprocess, the resultant fuzzy truth-value is converted to a crisp value. This can be done through the CENTROID method, which the crisp value of the output variable is computed by finding the variable value of the center of gravity of the membership function for the fuzzy value. At the same time, the MAXIMUM method returns the variable value at which the maximum truth-value occurs.

2.5.4 Applications of Expert Systems

The potential applications of Expert Systems for geographical problems were first pointed out by Smith (1984). Fisher *et al.* (1988) later reviewed several applications of Expert Systems in the geoscience field. However, generally speaking, geographical

Expert Systems are not well developed compared with those in the engineering and medicine fields. Some of these geographical Expert Systems are reviewed below.

Expert Systems are not uncommon in Remote Sensing and GIS applications (e.g., those reviewed by Robinson & Frank, 1987; Robinson *et al.*, 1987; Goodenough *et al.*, 1987). Expert System technology is also useful in geographic data handling, and resource and environmental management (e.g., Ripple & Ulshoefer, 1987; Davis *et al.*, 1986; Lein, 1993; Zeng & Zhou, 2001). Meanwhile, Expert Systems have been used for land cover classifications using Remote Sensing and GIS data (e.g., Wharton, 1987; Mason *et al.*, 1988; Bolstad & Lillesand, 1992; Kontoes *et al.*, 1993; Foschi & Smith, 1997; Bardossy & Samaniego, 2002). For example, an Expert System based on probability theory, later called Land Classification and Mapping Expert System (LCMES), was used for forest vegetation and soil classifications (Skidmore, 1989; Skidmore *et al.*, 1996). The Expert System handles uncertainty using probability theory, and consists of four main components; a knowledge base, a GIS, an inference engine, and a user interface. An obvious disadvantage of LCMES is that it does not have an explanation machine to help users understand the decision process.

Leung and Leung (1993) developed a typical Fuzzy Expert System shell for GIS, called Fuzzy-Logic-Based Expert System Shell (FLESS). The shell employs fuzzy logic for approximate reasoning and fuzzy set theory for handling uncertainty in database and rules base. It consists of four subsystems; a knowledge acquisition subsystem, a consultation driver, a fuzzy knowledge base, and a map display module. In the knowledge acquisition subsystem, facts can be obtained by asking users the values of the facts and the certainty of them, or they can be obtained from database or predefined files. The Left-Hand-Side of a rule can be a single proposition (evidence) or any combination of propositions connected either by logical *and* or logical *or*, while the Right-Hand-Side of a rule can only be a single proposition or multiple propositions with logical *and*. Meanwhile, certainty factors can be attached to each rule. The knowledge base is responsible for the storage and retrieval of fuzzy facts, fuzzy rules, and membership functions of fuzzy terms. Hashing techniques have been employed for efficient storage and retrieval. At the same time, the consultation driver includes an inference engine and an explanation machine. The inference engine supports both forward chaining and backward chaining. Evaluations of rules are based on fuzzy logic. The inference engine can also handle rules with multiple propositions and it uses an evidence combination approach for cases in which multiple rules have the same Right-Hand-Side. The explanation machine of FLESS is able to answer "why", "how", and "what if" questions. Furthermore, the map display module uses half-toning and coloring schemes for fuzzy displays.

Other applications of Fuzzy Expert Systems have been reported in Zhu *et al.* (1996), Penaloza & Welch (1996), Afonso *et al.* (1998), MacMillan *et al.* (2000), Zeng and Zhou (2001), and Bardossy and Samaniego (2002).

2.5.5 Problems of Expert Systems

One potential major problem of Expert Systems is the time and resource requirement of the traditional knowledge acquisition process. Interviewing domain experts and coding expert knowledge into computer programs usually costs much time and resources. Sometimes, the domain experts are not comfortable describing knowledge as rules. Also, domain experts do not always agree with each other. Another problem is how to effectively handle uncertainty in Expert Systems. It is believed that within many methods, fuzzy set theory is a good candidate in many cases. There are errors that may be involved in Expert Systems' building process, these include the expert's knowledge error, the semantic error associated with the knowledge acquisition process, the syntax error in the process of entering data and knowledge, the inference engine error, the inference chain error, and the error of ignorance (Girarratano & Riley, 1998). However, most of these errors can be corrected through careful planning and review in the process.

2.5.6 Summary of Expert Systems

Expert Systems are most suitable for ill-structured or non-structured domain problems. They are good at handling heuristic knowledge but at the same time, based on a logic reasoning foundation that is theoretically strong. There are other attractive features associated with Expert Systems such as flexibility, low cost, low risk, long-term benefit, combination of experts' knowledge, reliability, explanation capability, real time response, and intelligence (Basden, 1983; Girarratano & Riley, 1998). Nevertheless, Expert Systems, especially the traditional Expert Systems, have problems in acquiring expert knowledge and handling uncertainty.

Because of these limitations of Expert Systems, Basden (1983) argued that Expert Systems should not be applied to problems which are too simple, problems which are too complex, problems which require none of the advantages of Expert Systems, problems which rely on information more suited to processing by the human brain than by computer, and problems in "wide and shallow" domains. However, there are no simple rules that can be followed to determine whether or not the problem falls under the above conditions and therefore can be solved by Expert Systems.

2.6 COMBINATION OF MODELS

2.6.1 Principles of combination of models

A real world problem is usually so complex that no single model can do it altogether. Even though a single model can be applied to a problem, it may not be the best model for all aspects. Therefore, a combination of models becomes necessary and beneficial in many cases. The combination of models can be achieved at three levels. The lowest level – a combined model where sub-models are standalone and implemented individually, only the outcomes of these models are combined at the later stage. The second level – an integrated model consists of several sub-models each of which plays a distinct role either in data input or output or analysis processes. In the highest level – a hybrid model is where a subprocess of a model is replaced by another model to execute the function of the subprocess. The latter two models need user interfaces for smooth integration. Besides the combination of models at the three levels, there are standalone models with combined disciplines. Typical examples include such as Fuzzy Decision Trees (Chang & Pavlidis, 1977; Adamo, 1980; Janikow, 1998; Olaru and Wehenkel, 2003), Fuzzy Artificial Neural Networks (e.g., fuzzy ARTMAP used by Carpenter *et al.*, 1992; Carpenter *et al.*, 1997; Gopal *et al.*, 1999; Li *et al.*, 2002; Lin, 2004), Expert (rule-based) Neural Networks (Goodman *et al.*, 1992; Lacher *et al.*, 1992), as well as Fuzzy Expert Systems described in section 2.5.3.

2.6.2 Applications of combination of models

A typical hybrid model is, for example, where a Decision Tree or an Artificial Neural Network is used for the knowledge acquisition process of an Expert System. Huang and Jensen (1997) claimed that the machine learning models, in this case a C4.5 Decision Tree could be used for automated knowledge acquisition. They indicated that learning from examples was an effective and efficient alternative to the traditional knowledge acquisition method, because it escaped the "bottleneck" problem of the traditional knowledge acquisition process. For this purpose Decision Trees have long been seen as an easier way to build a knowledge base than Artificial Neural Networks. This is because a path in a Decision Tree can be naturally seen as a rule. But there may exist redundancy and contradiction if a path is transferred into a rule directly, therefore some kind of cleaning job must be done to delete redundancy and to resolve contradiction. In addition, two other examples have demonstrated the feasibility of combining Dempster-Shafer's theory and Expert Systems into hybrid models (Srinivasan & Richards, 1990; Peddle & Franklin, 1992).

An example of an integrated model was described by Lam and Pupp (1996), where the environmental information system RAISON consists of an Expert System, Environmental Models, an Artificial Neural Network, a GIS, and a database management system. The other integrated system is called IRMA (Integrated Resource Management Automation), and consists of a data base management system, a GIS, a rule-based Expert System, a data-exchange system, and an user interface shell (Loh & Rykiel, 1992).

In the lowest combination level, methods have been suggested to combine the results of standalone models. Xu *et al.* (1992) have reviewed some of these methods in a handwriting pattern recognition problem. They pointed out that there are three kinds of methods could be used to combine results of multi-classifiers either on an abstract level (i.e. output is a hard class label) or on a measurement level (i.e. output is probability measurements). The first one uses the majority voting principle, the second one uses a kind of a candidate subset combining and re-ranking approach, and the last one uses Dempster-Shafer's theory. Three such approaches have been eventually compared in some case studies. They are an averaged Bayes method, the majority voting methods, and Dempster-Shafer's theory. The results showed that the combination approaches achieved generally better performance than any of the individual classifiers. Among them the approach based on Dempster-Shafer's theory was more reliable and robust than others. One major assumption of their study is that individual classifiers are independent of each other. Their study has argued that there is a need to study the methodology of combining results of a number of different classification algorithms so that a better result could be obtained. In another study, Dempster-Shafer's theory was used to combine the results of several Artificial Neural Network classifiers in pattern recognition of hand-printed digits and letters (Rogova, 1994). The combination of classifiers decreased the misclassification error by 23% for digits, by 15% for uppercase letters, and by 25% for lowercase letters compared to the best individual classifier used in the combinations. Meanwhile, Rogova (1994) also found that it was important to find more independent classifiers to combine in order to achieve better performance. See *et al.* (1998) developed several combination approaches for hydrological modeling in two catchments. There were several individual models being used in the two catchments, these include Artificial Neural Networks, environmental models, models based on fuzzy set theory, and a naive model. Four combination approaches thus developed are a simply averaged method, the Bayes approach of Dougherty (1997) called Crisp Bayesian Model (CBM), a Fuzzy Bayes Model (FBM), and a Fuzzy Master Model (FMM). It was found that the fuzzy approaches (FBM and FMM) yielded better results than the other combination approaches and the individual models. They concluded that the combination approaches could give real advantages over individual models in handling large amounts of dynamic, non-linear or noisy data, and where there were underlying relationships which were not fully understood. Other advantages they identified included improved performance, faster model development and calculation times, and improved opportunities to provide estimates of prediction confidence through comprehensive bootstrapping operations. Besides those combination methods mentioned above, methods based on Stacked Regressions (Breiman, 1996; LeBlanc & Tibshirani, 1996), Discretization (Mojirsheibani, 1999), Correspondence Analysis (Merz, 1999), Production Rule (Steele, 2000), and convex combination (Carpenter *et al.*, 1997) have all been suggested for combining classifiers.

The other groups of combination classifiers needs to be mentioned is the so-called multiple classifiers systems (MCS). The common techniques of this group are boosting, bagging, consensus theory, and dynamic classifier selection (DCS), etc (McIver & Friedl, 2001; Smits, 2002; Briem *et al.*, 2002). These techniques create multiple classifiers by twisting a base classifier, and by combining them, classification performance may be improved. The intension of this study however, is to combine classifiers of different principles, which would potentially offer greater benefits.

2.6.3 Summary of combination of models

The combination of models is good for solving complicated environmental problems, when no single model can do the job satisfactory. The combination of models can occur at three levels. The lowest level – a combined model is more often encountered. Past studies showed that many combination methods could be used to combine individual models at this level, among which some are better than others. But most combined models were demonstrated to be better than any of the individual model, which certainly is encouraging. Another indication is that the combined model would be better if the individual models were based on very different principles (e.g., they are more independent). This is because those individual models that have very different principles tend to have different modes for handling the same problem, therefore by combining them they tend to correct one another and produce better modelling results. However, how to select appropriate individual models for the greatest combination benefit remains a problem (Petrakos *et al.*, 2001).

2.7 Summary

This chapter introduced the AI models and techniques used in this study, which include Artificial Neural Networks, Decision Trees, Dempster-Shafer's theory, fuzzy set theory, Expert Systems, and fuzzy rule-based expert systems. The application of these models and techniques to land cover classification and other geoscience areas has been reviewed. The chapter also discussed and summarized the problems and merits of each of these models and techniques.

In summary, these AI models and techniques are not parametric and have advantages over conventional statistical models. Artificial Neural Networks are the most theoretically complicated and problematic models among all these AI techniques. Many factors could significantly affect the Artificial Neural Network's modelling results. But

Artificial Neural Networks also show the greatest potential for environmental applications, as they still have much room to improve and develop in the AI field. Decision Trees are robust and simple models, but their application is limited by their suboptimal nature. Dempster-Shafer's theory should have received much more attention in the geoscience field, as it has advantages in dealing with incomplete data and hierarchical classification. Also, it can be used to handle uncertainty in Expert Systems. Fuzzy set theory is a great achievement of the 20^{th} century. It has pushed fundamental changes on control theory, decision-making theory, and scientific research. It is an appropriate tool to handle another source of uncertainty – fuzziness. It is strong on theory and versatile in application. The Expert System, however, is a mature AI technique. It is good at dealing with ill-structured problems and can give very sound comprehensibility. A Fuzzy expert system enjoys the advantages of both the traditional Expert System and fuzzy set theory.

Even though all of these AI techniques can be applied to land cover classification, none of them is the best overall. They all have their advantages and disadvantages. However, by combining several of these AI models, better classification performance may be achieved, as the combined model may cancel out problems and retain the merits of individual models. Combined models also save time and resources during an application. This is examined in Chapter 6. Prior to this the study area and data set are introduced in the next chapter.

Chapter 3

STUDY AREA, DATA AND PREVIOUS STUDIES

The chapter describes the geographical location and natural environment of the study area. The input data sources, which include seven independent variables and a dependent variable, and their quality, are first introduced. Then two data errors are identified. This is followed by the review of the previous studies at the study area.

3.1 DESCRIPTIONS OF THE STUDY AREA

The study area - Kioloa, is located on the south east coast of New South Wales, Australia (Figure 3.1). The area is extremely complex in both physiography and parent material (geology). This has resulted in a great variety of vegetation types with complex distributions from eucalypt-dominated sclerophyll forest to warm-temperate rain forest (Moore *et al.*, 1991).

For the forest area, there are about 450 species, 30 forest communities and 7 forest types. The aim of the study is to classify the area into 9 land cover classes including 7 forest types, a clear land / paddock class, and a water / sea class. The 7 forest types were aggregated from the 30 forest communities of Moore *et al.* (1991). They are: Dry Sclerophyll, E. *botryoides*, Lower slope wet forest, Wet E. *maculata*, Dry E. *maculata*, Rainforest Ecotone, and Rainforest. The boundaries of these 7 forest types are not clear-cut. They are separated on the basis of dominant species and the composition of understorey species. Another factor that complicates the classification task are disturbances such as fire and clearing.

3.2 DATA SOURCES AND THEIR QUALITY

There are seven independent input variables used in this study, which can be grouped into three sets. The first set is three bands of Landsat TM (Figure 3.2), in which band 2 ranges from 14-101 DN, band 4 ranges from 5-89 DN, and band 7 ranges from 0-86

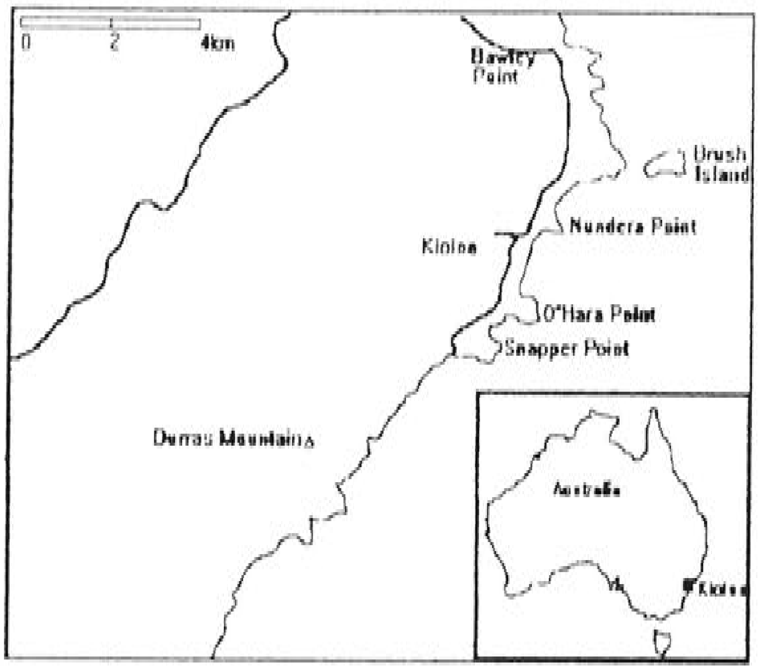


Figure 3.1 The location of the study area

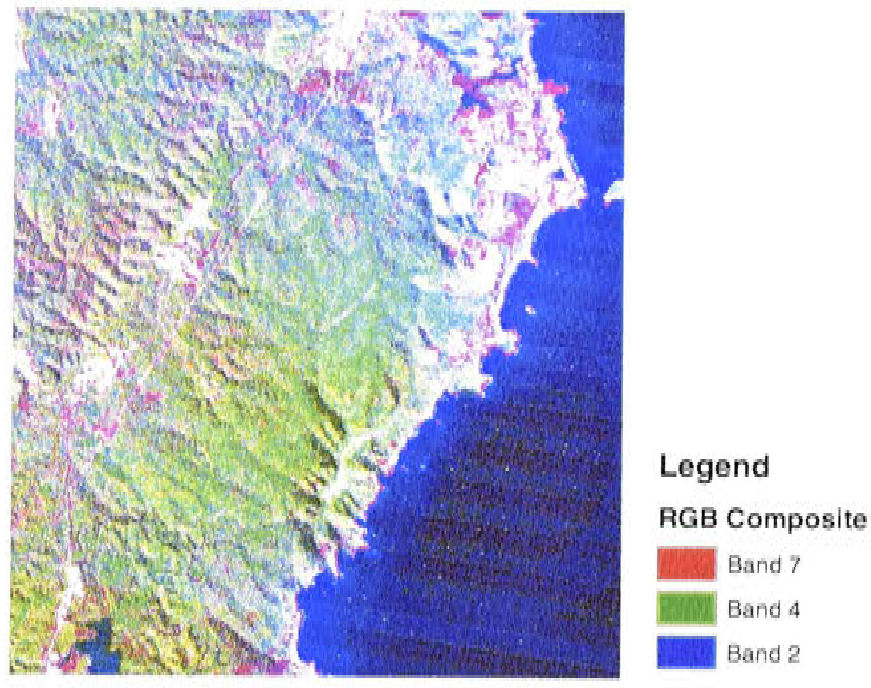


Figure 3.2 Composition of Band2, Band4 and Band7

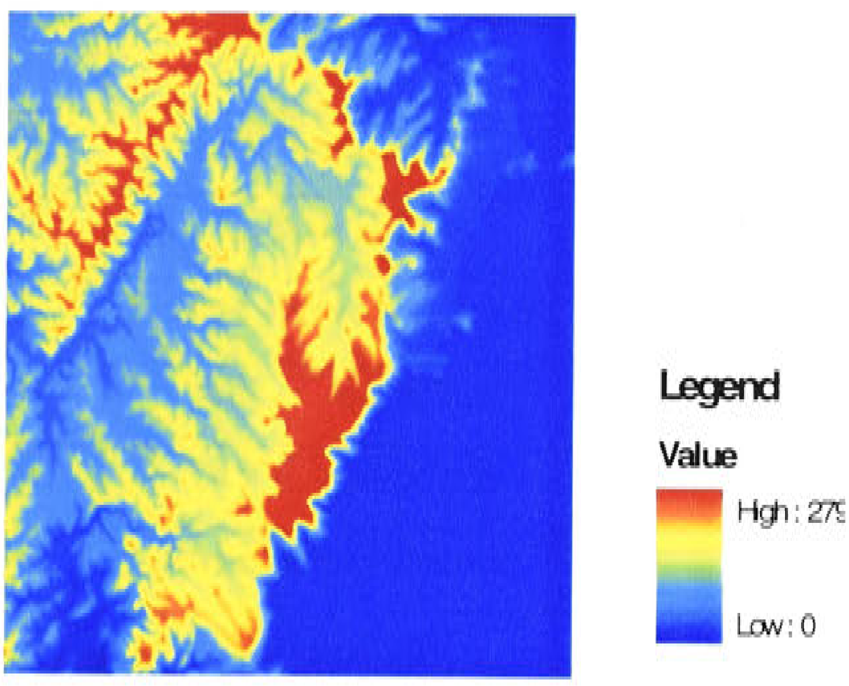


Figure 3.3 Digital Layer of DTM

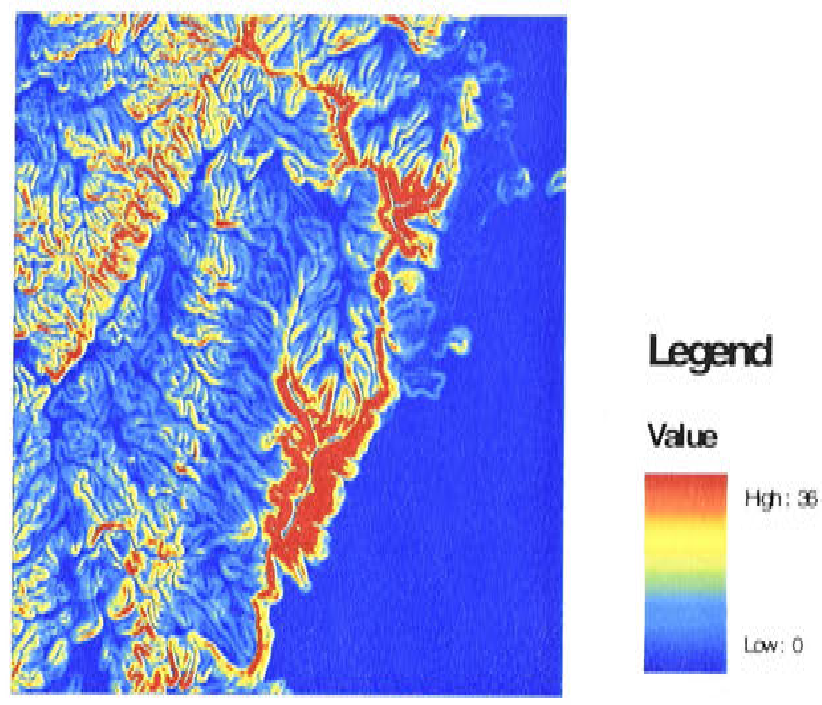


Figure 3.4 Digital Layer of Slope

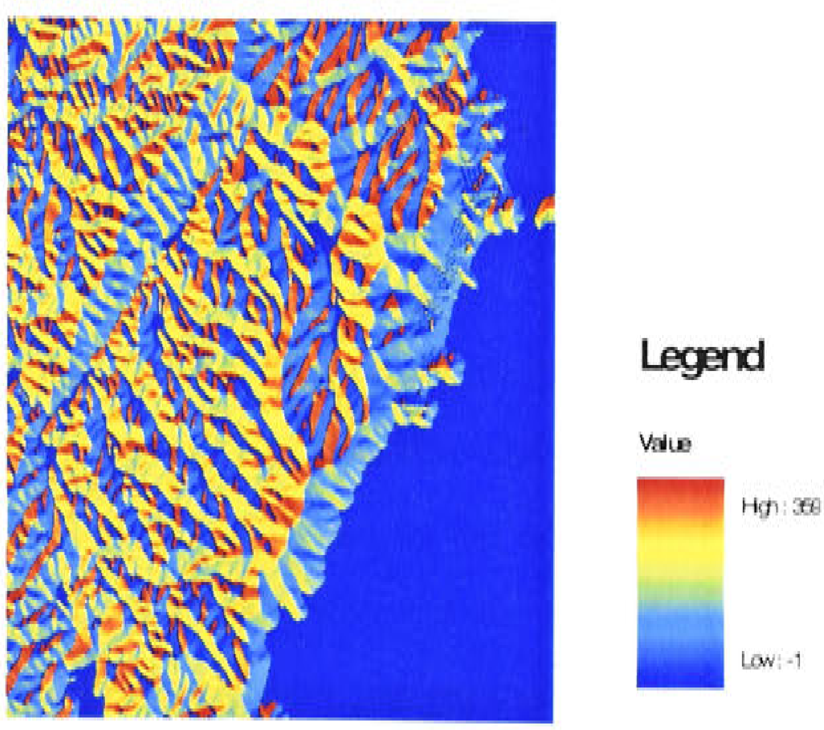


Figure 3.5 Digital Layer of Aspect

Figure 3.6 Digital Layer of Geology

DN. The second set is elevation (DTM) and its derivatives; slope and aspect, in which elevation ranges from 0-280 metres (Figure 3.3), slope ranges from 0-36 degree (Figure 3.4), and aspect ranges from –1-359 degree, where "aspect = -1" indicates flat ground (Figure 3.5). The third set is the geology variable (Figure 3.6), which has seven categories; Quaternary Alluvium, Tertiary Essexite, Snapper Point Permian, Pebbly Beach Permian, Wasp Head Permian, Ordovician, and Sea. Among the seven input independent variables, Landsat TM bands, DTM, and slope are interval data. The geology variable is categorical (nominal) data. The aspect variable is not interval data but a cyclic data type, as it represents directions rather than quantities. All of these seven input independent variables are stored as GIS (Arc/Info) grids with 30m spatial resolutions. The dependent variable is the field samples of the 9 land cover classes (Table 3.1).

Table 3.1: The number of samples of each class in the dataset

Class	Forest Type	Number of samples	Number of training samples	Numbers of test samples
1	Dry Sclerophyll	303	241	62
2	*E. botryoides*	69	48	21
3	Lower slope wet forest	52	37	15
4	Wet *E. maculata*	255	208	47
5	Dry *E. maculata*	180	137	43
6	Rainforest Ecotone	99	79	20
7	Rainforest	85	65	20
8	Cleared land/Paddock	166	136	30
9	Water/Sea	499	410	89

Within the total of 1708 field samples, 80% of them are randomly selected as the training set, and remaining 20% as the test set (Table 3.1).

Though there is documentation for the datasets (Fitzgerald, 1994), there is no data quality report. The Landsat TM data was acquired in April 1988, close to the survey of vegetation sites. It should be noted that the quality of the remotely sensed data could have been greatly affected by such factors as atmospheric effects, geometric aspects, sensor systems, data pre-processing, and so on (Lunetta, 1991). The DTM layer was digitised from the 1:25,000 topographic map of the study area therefore it was exposed to digitising error and errors inherent in the original map. The aspect and slope layers were derived from DTM using IDRISI's SURFACE function (Fitzgerald, 1994). While it is well known that there exist several different algorithms for deriving slope and aspect, the data error can be accumulated by any of these processes (e.g., Lee *et al.*, 1992). The digital geology layer was digitised from Gostin's (1969) 1:25,000 geological map, which also was exposed to digitising error and errors inherent in the original map.

Examining the air photos and Landsat image of the study area, three large geographical objects are most notable, one is Willinga River near the northeast corner of the land, another is Brush Island in the northeast of the sea, and the other is Durras Lake to the southwest (Figure 3.7). Also seen in the air photos and Landsat image is a power line easement which runs across the study area from southwest to northeast (Figure 3.7). Among the seven input variables, only the geology variable does not distinguish the island from the surrounding water. This is the first of the two serious data errors identified. Meanwhile, the ground samples were collected through several stages over four years by several authors (Fitzgerald, 1994). The samples are error-prone, not only because of the imbalance of sample representations, but also because of possible identification error and location error. After carefully checking the sample points with ground truth it was found that four samples located on Brush Island mentioned above have been labelled as the Water/Sea class rather than the correct land classes (probably the Clear land / Paddock class). This is the second serious data error, particularly because supervised classification requires good training samples for effective learning. The sampling error could have been corrected, but it was not, because one objective of this study is to look at how well AI models would handle the sampling error through the classification process.

3.3 PREVIOUS STUDIES AT THE STUDY AREA

The study area has been the subject of intensive research. Several Decision Trees and Artificial Neural Networks have been implemented for predictive forest mapping, as well as the Maximum Likelihood Classifier (e.g., Moore *et al.*, 1991; Lees and Ritman, 1991; Fitzgerald and Lees, 1992; Fitzgerald and Lees, 1994; Fitzgerald and Lees, 1996; Gahegan and West, 1998; German, 1999; Gahegan and Takatsuka, 1999; and Lees, 1996a; 1996b; 1996c). Reviews of these past studies can help gain insights into the datasets, the application, and the models.

Moore *et al.* (1991) described a Decision Tree - CART (Breiman, *et al.*, 1984) used for predicting the 30 forest communities in the forested part of the study area. Only environmental variables were used; these included geology, and topographic variables of slope, azimuth, aspect, horizon, downhill position, uphill position, catchment, steepness (local roughness), and watershed. The predictive model was based on the association between the distribution of vegetation and many environmental variables, among which elevation and geology have special significance. A total of 90% of the forest samples (1257) was used for training the Decision Tree. The Decision Tree was then applied to the remaining area. It achieved an overall accuracy of 83% over 10% test samples. Although not calculated at the time, levels of confidence were low. But this application confirmed the value of the Decision Tree model and GIS environmental modeling for mapping floristic forest classes.

Subsequently, Lees and Ritman (1991) added remotely sensed data (Landsat TM) to the database to include some disturbance information. They claimed that the integrated approach using both the remotely sensed data and the environmental variables offered the possibility of improving on the thematic mapping capabilities of both environmental modelling and the classification of remotely sensed data. In their study, Lees and Ritman (1991) trained CART to discriminate the 7 forest types plus a category of cleared land instead of the 30 forest communities. The results showed that only the Dry Sclerophyll forest and the cleared land were satisfactorily predicted, other forest types having accuracies ranging from 19% to 49%. The overall training accuracy was 56.8%. However, this low accuracy compared with that of the last study (83%) was understandable as Lees and Ritman (1991) included the disturbed and cleared areas around the forest core used in the earlier study, making the task much more complex.

In 1992, Fitzgerald and Lees applied a standard backpropagation Artificial Neural Network to the study area. The Artificial Neural Network includes an input layer of 8 nodes representing three bands of Landsat TM and five environmental variables, an output layer of 10 nodes representing 9 output classes of the 7 forest types, a class of cleared land, and a class of water, and a hidden layer of 24 nodes. After training the network, they achieved an overall training accuracy of 51% for the land portion. Compared with the results of Lees and Ritman (1991) who applied a Decision Tree, the floristic patterning was more sophisticated and less polygonal.

Fitzgerald and Lees (1994) went on to report the results of the implementation of a range of spatial context scales into the Artificial Neural Network. Context windows of 3x3, 5x5, 7x7, and 9x9 were fed into the input nodes with multisource datasets, and the network was trained with different numbers of hidden nodes. The resulting accuracies were better than the Maximum Likelihood Classifier (training accuracy of 50.5%), the Decision Tree (training accuracy of 56.8%), and the Artificial Neural Network with no spatial context (45.7%). They concluded that different spatial contexts did indeed produce better mapping accuracies for each forest type. However the large sample sizes necessary were unlikely to be cost-effective in practical application. Temporal context, on the other hand, could also improve the prediction accuracy (Fitzgerald and Lees, 1996), and this was more cost-effective.

In 1998, Gahegan and West applied Artificial Neural Networks and Decision Trees to the Kioloa dataset. They reported that DONNET (Discrete Output Neural NETwork) performed about 10% (test accuracy around 70-75%) better than C4.5, and they found that both methods were much better than the Maximum Likelihood Classifier which produced only 40% accuracy. German (1999) has used the DONNET, C4.5 and Vanilla (straightforward) back-propagation on the same dataset, and they obtained overall accuracies of 72.61%, 66.96%, and 50.77% from validation samples. He has shown further improvement of the performance of DONNET by restricting the movement of the hyperplanes. The freezing of the appropriate weights and recalculation of the hidden layer/output layer weights achieved an accuracy of 75.4%.

Subsequently, Gahegan and Takatsuka (1999) tested the combination of an unsupervised Artificial Neural Network and a supervised Artificial Neural Network. Combining a Self-Organising Map (SOM) and a Learning Vector Quantisation (LVQ) increased overall accuracy from 66.45% to 68.63% on the Kioloa dataset, although earlier applications of the LVQ by Lees gave much poorer results (Lees, 1996d). The difference appeared to be that normalizing the input data resulted in improved performance.

3.4 Summary

The chapter shows that the study area is an extremely complex natural environment with forest types changing gradually across a continuum. Breaking the continuum into the seven discrete forest types poses great difficulty for the classification task. This is further complicated by disturbance from fire and clearing.

The multisource GIS data used for this study have various data types including interval data, categorical data, and mixed type data. They are open to instrument and human errors. Two data errors were identified by scrutinizing the data. One is sampling error, the other is attribute error associated with the geology variable. Obviously, this error-prone data will affect the performance of the forest type mapping.

Reviews of past studies show that several Decision Trees and Artificial Neural Networks as well as traditional statistical models such as the Maximum Likelihood Classifier have been implemented for predictive forest mapping. Lees (1996a) reported that the best results of output from the backpropagation Artificial Neural Network, Decision Trees and the Maximum Likelihood Classifier are similar and concluded that it was the data model itself that limits the improvement of predictive accuracy, not the classifiers. Several studies, however, found much better predictive accuracies by using Artificial Neural Networks other than the backpropagation neural network. These Artificial Neural Networks can avoid some problems of the standard backpropagation neural network, but they also increase the complexity for the implementation.

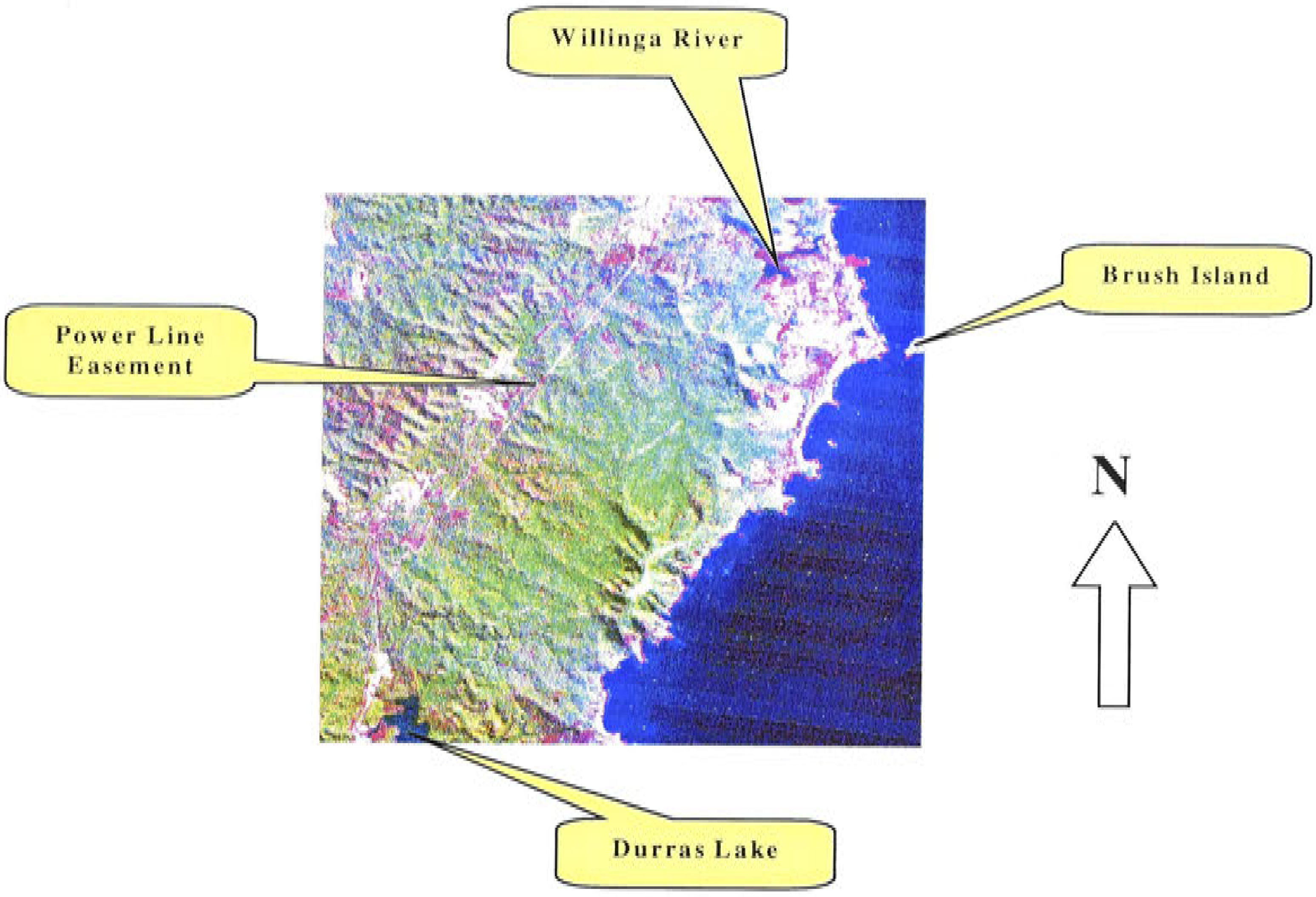


Figure 3.7: Locations of several geographical objects

Chapter 4

CONCEPTUAL MODEL AND ACCURACY ASSESSMENT METHODS

This chapter describes the conceptual model of this study. The conceptual model indicates that there were four modeling stages in this study, each of which focused on one topic. This chapter also describes the accuracy assessment methods used in this study for complicated forest type mapping.

4.1 CONCEPTUAL MODEL

The conceptual model of this study is displayed on the flow chart of Figure 4.1. It indicates four stages of this study. In the first stage, three individual AI models including a Decision Tree (DT), an Artificial Neural Network (ANN), and a model based on Dempster-Shafer's theory (D-S) used the multisource GIS data (see section 3.2) to separately classify the study area into the 9 land cover classes (e.g., represented by the blue dashed arrow on the flow chart). The purpose was to see whether or not the three individual AI models were capable classifiers for complicated forest type mapping and what advantages and disadvantages each of them has in the mapping process. In the second stage, the results of the three individual AI models were combined using different approaches for the complicated forest type mapping (e.g., represented by the red dashed arrow on the flow chart). The purpose was to examine the advantages of the combination strategy over individual models for forest type mapping. As two obvious data errors were identified from the database, the third stage was to evaluate the modes of the three individual AI models and the effectiveness of the combined AI models in handling the data errors (e.g., represented by the green dashed arrows on the flow chart). In the fourth stage, several Fuzzy Expert Systems (ES) were built by learning directly from samples that were selected either from the results of the combined AI models or from the field survey (e.g., represented by the yellow dashed arrow on the flow chart). These Fuzzy Expert Systems were then applied to complicated forest type mapping and compared to the three individual AI models used in the first stage of this study. Detailed descriptions of these methods and approaches are presented in the following chapters.

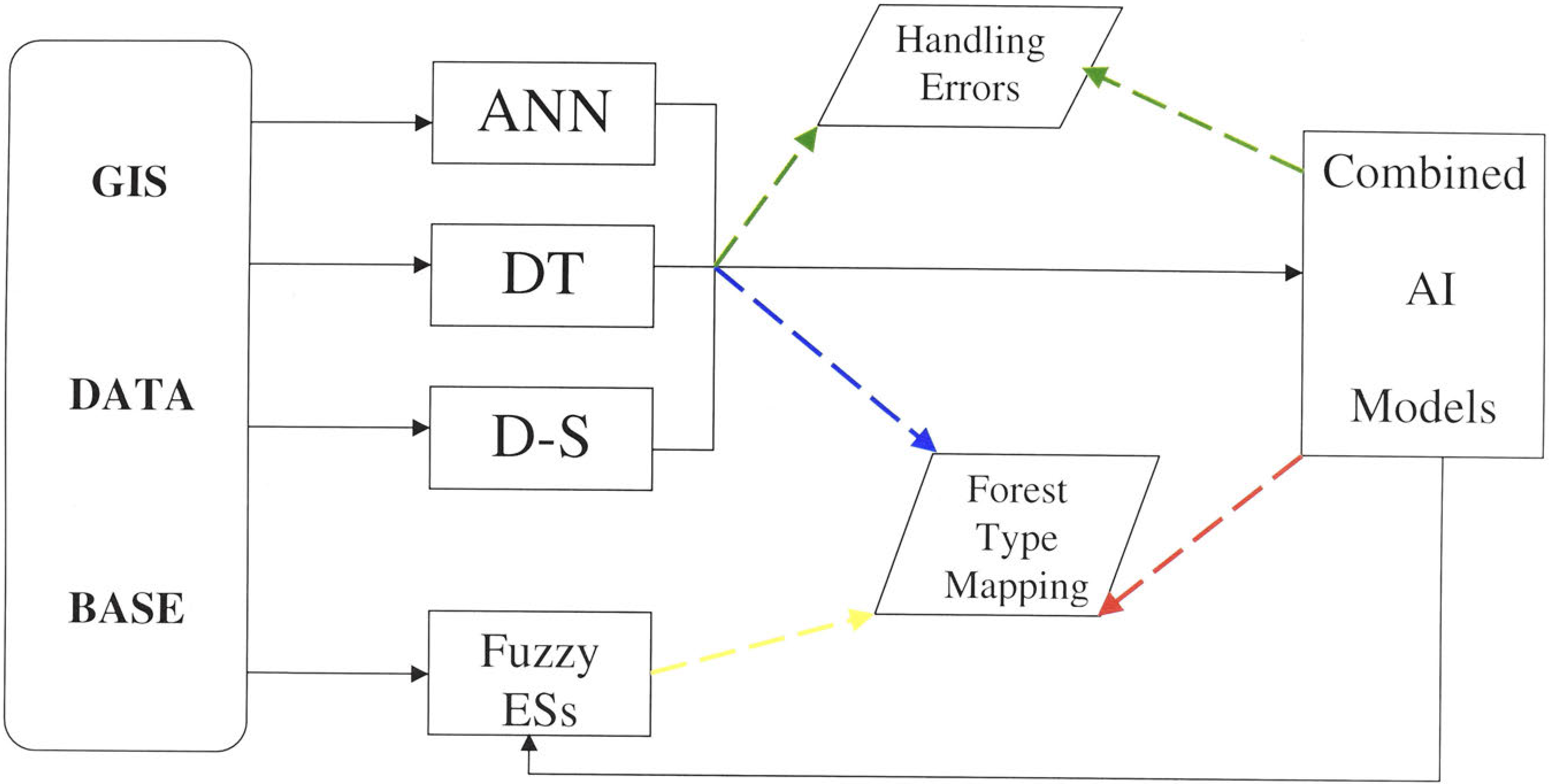


Figure 4.1 Flow chart of this study

4.2 ACCURACY ASSESSMENT METHODS

Traditionally, an error or confusion matrix (contingency table) is often used for assessing classifications. From the error matrix, overall accuracy, producer's accuracy, user's accuracy, Kappa accuracy, and several other accuracy measurements can be obtained (e.g., Story & Congalton, 1986; Cohen, 1960; Ma & Redmond, 1995; Prisley & Smith, 1987; Rosenfield & Fitzpatrick-Lins, 1986; Congalton *et al.*, 1983; Congalton, 1991). For example, the following is an error matrix (Table 4.1):

Table 4.1 An example of classification error matrix

Classified Data	**Reference Data**					
	j = 1	.	...	.	j = k	Total
i = 1	X_{11}	.	...	.	X_{1k}	X_{1+}
.	.	.	...	.	.	.
.	.	.	X_{ij}	.	.	X_{i+}
.	.	.	...	.	.	.
i = k	X_{k1}	.	...	.	X_{kk}	X_{k+}
Total	X_{+1}	.	X_{+j}	.	X_{+k}	X_{++}

X_{i+} = total for row i;
X_{+j} = total for column j;
X_{++} = total number of samples in error matrix;
X_{ij} = number of samples classified as class i which are found to belong to class j on the reference data;
k = number of classes;
i = row index; and
j = column index

Then overall accuracy is calculated as:

$$Overall\ Accuracy = \frac{\sum_{i=1}^{k} X_{ii}}{X_{++}} \tag{4.1}$$

Producer's accuracy is associated with the error of commission and relates to the probability that a reference sample will be correctly classified. It is calculated as:

$$Producer's\ Accuracy = \frac{X_{ii}}{X_{+i}} \tag{4.2}$$

User's accuracy is associated with the error of omission and relates to the probability that a pixel on the classification map was correctly classified. It is calculated as:

$$User's\ Accuracy = \frac{X_{ii}}{X_{i+}} \tag{4.3}$$

Kappa accuracy was first introduced by Cohen (1960) and has received wide acceptance throughout the Remote Sensing and GIS fields as a better standard for evaluating classification accuracy than the overall accuracy. It is calculated as:

$$K = \frac{P_o - P_c}{1 - P_c} \tag{4.4}$$

where P_o is predicted agreement, and it is calculated as:

$$P_o = \frac{\sum_{i=1}^{k} X_{ii}}{N} \tag{4.5}$$

where N is the total number of samples in the error matrix (e.g., N=X_{++}). While P_c is agreement predicted by chance, and it is calculated as:

$$P_c = \frac{\sum_{i=1}^{k} X_{i+} X_{+i}}{N^2} \tag{4.6}$$

Due to the identified sampling error associated with the Water/Sea class, which 4 samples were misclassified as the Water/Sea class possibly from the Clear land/Paddock class (see section 3.2), **the study has focused on assessing the predictive accuracies only for the 7 forest types**. With attention focused on the 7 forest types, producer's accuracy and user's accuracy can be calculated as usual. Overall accuracy can be calculated as below:

$$Overall\ Accuracy = \frac{\sum_{i=1}^{7} X_{ii}}{\sum_{i=1}^{7} X_{+i}} \tag{4.7}$$

To calculate Kappa accuracy for only the 7 forest types from an error matrix of total 9 classes, the above formulas for P_o and P_c must be modified to:

$$P_o = \frac{\sum_{i=1}^{k} X_{ii}}{N_1} \tag{4.8}$$

$$\text{and } P_c = \frac{\dfrac{\sum_{i=1}^{k} X_{i+} X_{+i}}{N}}{N_1} \tag{4.9}$$

where $N_1 = \sum_{i=1}^{7} X_{+i}$ is the total number of field samples for the seven forest types (228 in this case). The numerator of P_c represents the number of samples predicted by chance which are calculated from marginal probabilities (Cohen, 1960), while the numerator of P_o represents the number of samples which are actually predicted by the classifier.

In this study, all classifiers were assessed by using Equation 4.7 for overall accuracy, Equation 4.2 for producer's accuracy, Equation 4.3 for user's accuracy, and Equations 4.4, 4.8 and 4.9 for Kappa accuracy.

4.3 SUMMARY

This chapter describes the four modeling stages of this study shown in the conceptual model of Figure 4.1. The first stage used three individual AI models for complicated forest type mapping. The second stage combined the results of the three individual AI models for complicated forest type mapping. The third stage examined the modes of the individual AI models and the effectiveness of the combined AI models in handling data errors. The fourth stage built Fuzzy Expert Systems and applied them to complicated forest type mapping. The following four chapters focus on each of these modeling stages. Each chapter includes a description of methodology, a report of results and findings, a detailed discussion, and a summary.

This chapter also describes the accuracy assessment methods used throughout the study. They are overall accuracy, producer's accuracy, user's accuracy, and Kappa accuracy. Modifications were made for assessing classification accuracy of the 7 forest types from an error matrix of 9 land cover classes.

Chapter 5

INDIVIDUAL AI MODELS FOR PREDICTIVE FOREST TYPE MAPPING

The chapter presents the first stage of this study, in which three individual AI models including a Decision Tree (DT), an Artificial Neural Network (ANN), and a model based on Dempster-Shafer's theory (D-S) were applied to complicated predictive forest type mapping using multisource GIS data. First, the chapter describes the three individual AI models in detail. Then, the chapter reports the results of the three classifications in terms of predictive accuracies and visual appearance. Following that, the chapter discusses the results and findings. Finally, the chapter gives a short summary of this stage of study.

5.1 METHODS

This study chose a Decision Tree, an Artificial Neural Network, and a model based on Dempster-Shafer's theory for complicated predictive forest type mapping. The three AI models are non-parametric, and they have advantages over conventional statistical models such as the Maximum Likelihood Classifier. The three AI models make no assumption about the data distribution, and therefore have avoided a possible error source. Also, the three AI models can readily use multisource data without additional effort. Past studies have shown that the three AI models were generally better classifiers than conventional parametric models, especially in complicated classification problems (see sections 2.1-3). In addition, the three AI models were said to be more error tolerant than parametric models (Quinlan, 1986; Mingers, 1989; Hepner *et al.*, 1990; Moon, 1990; see sections 2.1-3).

5.1.1 Decision Tree

This study employed a Decision Tree provided by the S-Plus package, which is developed from the CART algorithm (Clark & Pregibon, 1992). A Split Number of 15 was selected, which means that each decision node at least has 15 training samples to split, a node with less than 15 training samples will be left as a terminal node or leaf.

There was no pruning algorithm implemented in this study. The Decision Tree accepted continuous and categorical data and produced both hard classification and soft probability outcomes.

5.1.2 Artificial Neural Network

The Artificial Neural Network used in this study is a standard multi-layer feed-forward Artificial Neural Network trained using the backpropagation algorithm provided by the SNNS package (Stuttgart Neural Network Simulator) (Zell *et al.*, 1995). The architecture of the Artificial Neural Network includes an input layer of 8 nodes to represent 8 input variables (note: the aspect variable was split into two variables; cosine of aspect and sine of aspect), a hidden layer of 10 nodes, and an output layer of 9 nodes to represent the 9 output classes. The parameters are 0.4 for step size, 0.2 for momentum, 0.1 for flat spot elimination, and 0.1 for error tolerance. To facilitate the Artificial Neural Network, the input variables were normalized into values between 0 and 1. The training started from random weights and stopped until the error tolerance was met or after 100 cycles, whatever comes first. The trained Artificial Neural Network was then applied to classify the whole study area. It should be noted that the Artificial Neural Network was run only once, and no fine-tuning or parameter optimization process has been implemented. This is because one objective of this study is to see whether or not applying combined AI models can improve the classification performance without spending time optimizing the Artificial Neural Network. The Artificial Neural Network also produced both hard classification and soft probability outcomes.

5.1.3 Dempster-Shafer's Theory of Evidence

In this study an approximation of Dempster's rule of combination was applied to the forest type mapping in order to avoid the problems that may arise when there is zero or one assigned to the *mass of support* from independent evidence sources. For example, let us assume that the target classes are exhaustive exclusive, i.e., c_1, c_2, and c_3 do not overlap, and any pixel can be assigned to only one of the three classes. It should be noted that this is assumed to be valid for this study and through the whole thesis. Now, the masses of support assigned from five independent sources are listed as follow:

1. source1 : $m(c_1) = 0$, $m(c_2) = 1$, and $m(c_3) = 0$

2. source2 : $m(c_1) = 0.2$, $m(c_2) = 0.4$, and $m(c_3) = 0.4$
3. source3 : $m(c_1) = 0.4$, $m(c_2) = 0.3$, and $m(c_3) = 0.3$
4. source4 : $m(c_1) = 0$, $m(c_2) = 0$, and $m(c_3) = 1$
5. source5 : $m(c_1) = 0.5$, $m(c_2) = 0.1$, and $m(c_3) = 0.4$

The question is what are the updated masses of support for the three classes after applying Dempster's rule of combination?

If we combine the five evidence sources following the order of; source1, source2, source3, source4, and source5, the masses of support after combination are; $m(c_1) = 0.5$, $m(c_2) = 0.1$, and $m(c_3) = 0.4$. The general formula of Dempster's rule of combination is:

$$m_1 \oplus m_2(Z) = \frac{1}{1-k} \sum_{X \cap Y = Z} m_1(X) m_2(Y)$$

$$\text{where } k = \sum_{X \cap Y = \phi} m_1(X) m_2(Y) \tag{5.1}$$

where m_1 and m_2 are the masses of support from two items of evidence, X and Y are the elements of Θ (*Frame of discernment*), $Z = X \cap Y$, and k is the mass that the combination assigns to the null subset $\emptyset$. For example,

- The resultant mass table for the first combination (i.e., source1 + source2) is Table 5.1:

Table 5.1: The mass table for the first combination

source1\source2	**C1/0.2**	**C2/0.4**	**C3/0.4**	**Θ/0**
C1/0	C1/0	Ø/0	Ø/0	C1/0
C2/1	Ø/0	C2/0.4	Ø/0	C2/0
C3/0	Ø/0	Ø/0	C3/0	C3/0
Θ/0	C1/0	C2/0	C3/0	Θ/0
Normalised masses	C1/0	C2/1	C3/0	Θ/0

- The resultant mass table for the second combination (i.e., source1 + source2 + source3) is Table 5.2:

Table 5.2: The mass table for the second combination

1st combination\source3	C1/0.4	C2/0.3	C3/0.3	Θ/0
C1/0	C1/0	Ø/0	Ø/0	C1/0
C2/1	Ø/0	C2/0.3	Ø/0	C2/0
C3/0	Ø/0	Ø/0	C3/0	C3/0
Θ/0	C1/0	C2/0	C3/0	Θ/0
Normalised masses	C1/0	C2/1	C3/0	Θ/0

- The resultant mass table for the third combination (i.e., source1 + source2 + source3 + source 4) is Table 5.3:

Table 5.3: The mass table for the third combination

2nd combination\source4	C1/0	C2/0	C3/1	Θ/0
C1/0	C1/0	Ø/0	Ø/0	C1/0
C2/1	Ø/0	C2/0	Ø/0	C2/0
C3/0	Ø/0	Ø/0	C3/0	C3/0
Θ/0	C1/0	C2/0	C3/0	Θ/0
Normalised masses	C1/0	C2/0	C3/0	Θ/1

- The resultant mass table for the fourth combination (i.e., source1 + source2 + source3 + source4 + source5) is Table 5.4:

Table 5.4: The mass table for the fourth combination

3rd combination\ ource5	C1/0.5	C2/0.1	C3/0.4	Θ/0
C1/0	C1/0	Ø/0	Ø/0	C1/0
C2/0	Ø/0	C2/0	Ø/0	C2/0
C3/0	Ø/0	Ø/0	C3/0	C3/0
***Θ*/1**	C1/0.5	C2/0.1	C3/0.4	Θ/0
Normalised masses	C1/0.5	C2/0.1	C3/0.4	Θ/0

In the first combination, information from source2 is completely masked by source1 because source1 has zero and one in the masses of support. The same occurs in the second combination. In the third combination however, a contradiction arises when 100% belief is assigned to two different classes. Dempster's rule of combination is unable to resolve this contradiction, but simply assigns zero belief to all three classes, which causes total *ignorance* (i.e., $m(\Theta) = 1$) completely eroding the combined information of the last four sources. Thus source5 is the only evidence left to provide masses of support in the fourth combination. So, if the combined order is; source1, source3, source5, source4, and source2, then the results are; $m(c_1) = 0.2$, $m(c_2) = 0.4$, and $m(c_3) = 0.4$. This indicates that combining the evidence sources in a different order produces different outcomes. This violates one principle of Dempster's rule of combination which is that the order of combination will not change the final result. From the above analysis, it is shown that Dempster's rule of combination could not resolve a contradiction, and that the combination order generates an error when zero and one are assigned to masses of support. To avoid this problem one should combine contradictory evidences (i.e., source1 and source4) first and, instead of assigning zero to each class, divide the total ignorance into the two contradictory classes. For example, $m(c_1) = 0$, $m(c_2) = 0.5$, and $m(c_3) = 0.5$. The final results following this revised approach are; $m(c_1) = 0$, $m(c_2) = 0.2$, and $m(c_3) = 0.8$.

The essential part of the above revised approach is to identify the correct order in which to combine the evidences and to find out which two items of evidence are contradictory, in order to resolve this. For a real, large database this is difficult, if not impossible. Practically, we chose to resolve this problem by using an approximation of Dempster's rule of combination. Instead of assigning 1 to one class, and 0 to all others, we assign a number smaller than 1, such as 0.99 to the former class, and a number close to 0, such as 0.01, to the latter classes. Meanwhile, instead of using the formula (5.1) we use:

$$m_1 \oplus m_2(Z) = \frac{1}{k} \sum_{X \cap Y = Z} m_1(X) m_2(Y)$$
$$\text{where } k = \sum_{X \cap Y \neq \varphi} m_1(X) m_2(Y) \qquad (5.2)$$

The formula 5.2 results in much fewer calculations and therefore faster implementation when output classes are exhaustive exclusive.

This approximation of Dempster's rule of combination resolves the contradiction and escapes the ordering effect. For example, if we assign $m(c_1) = 0.01$, $m(c_2) = 0.99$, and $m(c_3) = 0.01$ from the source1, and assign $m(c_1) = 0.01$, $m(c_2) = 0.01$, and $m(c_3) = 0.99$ from the source4. If we combine the sources in the order; source1, source2, source3, source4, and source5, the updated masses of support are; $m(c_1) = 0.006$, $m(c_2) = 0.198$, and $m(c_3) = 0.796$ which is very close to the correct result, and the result would not change with the changed order. It should be noted that this approximation does not require the masses of support be summed to 1, but a value very close to 1 is sufficient.

This study applied the above-developed approximation of Dempster's rule of combination to combine masses of support from the following independent data sources. The three Landsat TM bands were combined as one independent data source because of their similar characteristics. DTM, slope, aspect, and geology were treated as another four independent data sources. The aim was to use the training samples to obtain the mass of support for each class at each pixel from each of these five independent data sources. For Landsat TM, DTM, and slope, because they are interval data, the following command and functions of Arc/Info Grid were used to calculate the masses of support (probabilities) of each pixel belong to the output classes: "MAKESTACK", "CLASSSIG()", and "CLASSPROB()". For example,

1. MAKESTACK slopes list slope,
2. slope.gsg = CLASSSIG(slopes, samtrg), and
3. slopeprob = CLASSPROB(slopes, slope.gsg),

where "slope.gsg" is a signature file and "samtrg" is a grid of the training samples. The CLASSPROB function outputted a stack of grids (9 grids to represent the 9 output classes). There was one grid for each class in the input signature file. Each layer in the stack (i.e., "slopeprob") stored the probability that a pixel belongs to that class.

Because the geology variable is categorical data, the above method could not be used. Instead, a method has been developed for this study to derive the masses of support from the geology variable, which is described next. Firstly, an "area matrix" (Table 5.5) between the training samples (samtrg) and the geology variable was obtained in ArcView using the menu of "Tabulate Areas". For example, the obtained matrix is:

Table 5.5 Area matrix between samtrg and geology

samtrg / geology	Type1	Type2	Type3	Type4	Type5	Type6	Type7	Sum
Class1	0	0	0	10800	70200	21600	114300	216900
Class2	0	0	0	6300	24300	0	12600	43200
Class3	0	0	0	2700	13500	0	17100	33300
Class4	0	0	900	27000	28800	1800	128700	187200
Class5	0	0	0	29700	45000	1800	46800	123300
Class6	0	0	0	1800	4500	900	63900	71100
Class7	0	0	0	23400	9900	0	25200	58500
Class8	1800	12600	46800	22500	5400	0	33300	122400
Class9	369000	0	0	0	0	0	0	369000

Note: the values in the table represent area, in which a value of 900 represents the area of a single pixel with 30 metre spatial resolution

Secondly, the value of each cell of Table 5.5 was divided to its row sum. For example, the resultant matrix (Table 5.6) is:

Table 5.6 Matrix resulting from dividing the value of each cell of table 5.5 by its row sum

samtrg / geology	Type1	Type2	Type3	Type4	Type5	Type6	Type7
Class1	0	0	0	0.049793	0.323651	0.099585	0.526971
Class2	0	0	0	0.145833	0.5625	0	0.291667
Class3	0	0	0	0.081081	0.405405	0	0.513514
Class4	0	0	0.004808	0.144231	0.153846	0.009615	0.6875
Class5	0	0	0	0.240876	0.364964	0.014599	0.379562
Class6	0	0	0	0.025316	0.063291	0.012658	0.898734
Class7	0	0	0	0.4	0.169231	0	0.430769
Class8	0.014706	0.102941	0.382353	0.183824	0.044118	0	0.272059
Class9	1	0	0	0	0	0	0
Sum	1.014706	0.102941	0.387161	1.270954	2.087006	0.136457	4.000775

Then, the value of each cell of Table 5.6 was divided by its column sum to normalize it. For example, the resultant matrix (Table 5.7) is:

Table 5.7 Matrix resulting from dividing the value of each cell of table 5.6 by its column sum

samtrg / geology	Type1	Type2	Type3	Type4	Type5	Type6	Type7
Class1	0	0	0	0.039177	0.155079	0.72979	0.131717
Class2	0	0	0	0.114743	0.269525	0	0.072903
Class3	0	0	0	0.063795	0.194252	0	0.128353
Class4	0	0	0.012418	0.113482	0.073716	0.070464	0.171842
Class5	0	0	0	0.189524	0.174874	0.106983	0.094872
Class6	0	0	0	0.019919	0.030326	0.092763	0.22464
Class7	0	0	0	0.314724	0.081088	0	0.107671
Class8	0.014493	1	0.987582	0.144634	0.021139	0	0.068002
Class9	0.985507	0	0	0	0	0	0

The values in Table 5.7 indicate the masses of support derived from the geology variable using the training samples. The values in Table 5.7 were then transformed into nine grids in the Arc/Info Grid environment, each of which stored the probability that a pixel belongs to that class.

The aspect variable is not interval data but a cyclic data type and therefore it is inappropriate to apply the same method used for the Landsat TM bands, DTM, and slope. For this reason, the aspect variable was divided into nine categories, among which eight categories represent eight directions (e.g., north, northeast, east, southeast, south, southwest, west, and northwest) and the last one represents the flat area which has "aspect = -1". Then, the same method used for the geology variable was applied to derive the masses of support for the aspect variable using the training samples.

After having derived the masses of support from the five independent sources, the approximation of Dempster's rule of combination described above were used to combine them, and this eventually produced the probability classification outcomes of the study area. The modeling process was done by writing an AML (Arc/Info Macro Language) and by executing it in the Arc/Info Grid environment.

5.2 RESULTS

For the Decision Tree, the resultant tree structure is displayed in Figure 5.1. It can be seen that the geology variable appears at the top level (i.e., root) of the tree, and class 9 (Water/Sea) is determined only by geology type 1 (Sea).

For the Artificial Neural Network, Figure 5.2 shows the plot of training MSE (Mean Squared Error) and testing MSE through the modeling process. It indicates that after 100 cycles of training, both the training MSE and the testing MSE still could not reach the desired error tolerance of 0.1. The training MSE decreased sharply in the first few cycles then reduced smoothly. The testing MSE became steady after about 50 cycles. The network behavior at one point of time was captured in Figure 5.3. The network architecture shows a three-layer feed-forward Artificial Neural Network with connections between layers. The size of each neuron represents the output value of the neuron (e.g., the number below the neuron). A proportion of the network weights have also been displayed on the figure.

For the model based on Dempster-Shafer's theory, a correlation matrix was first calculated to test the assumption of independence (Table 5.8). It indicates that there is not high correlation between input variables; but there is a moderate correlation between band 7 and band 2, band 7 and band 4, geology and aspect, geology and band 4, DTM and band 4, DTM and slope, as well as DTM and geology. Therefore the assumption of evidence independence is not well satisfied. However, as argued by Lee *et al.* (1987), the advantages associated with the assumption of independence outweigh the disadvantages. More importantly, violating the assumption makes it possible to test the uncertainty tolerance of the combination approaches.

The error matrix of the Decision Tree is shown in Table 5.9. The error matrix of the Artificial Neural Network is shown in Table 5.10. Table 5.11, meanwhile, shows the resultant error matrix of the model based on Dempster-Shafer's theory. Among the three individual AI models, the Decision Tree achieved the best overall and Kappa accuracies over the 7 forest types, followed by the Artificial Neural Network and the model based on Dempster-Shafer's theory. The error matrices reveal that no single model has the best user's accuracies and producer's accuracies for all of the 7 forest types. This created the possibility of performance improvement using the combination strategy.

Comparing the three classification maps of the Decision Tree (Plate 1), the Artificial Neural Network (Plate 2), and the model based on Dempster-Shafer's theory (Plate 3) shows that they look very different in appearance. This is not unexpected because the three models are based on different principles and have very different characteristics in predicting forest types. For example, the model based on Dempster-Shafer's theory and the Artificial Neural Network have correctly predicted Durras Lake near the southwest corner of the study area (see Figure 3.7), but the Decision Tree has completely misclassified the area. On the other hand, while the Decision Tree has perfectly predicted Willinga River near the northeast corner (see Figure 3.7), the model based on Dempster-Shafer's theory has only predicted it partly, and the Artificial Neural Network is worse. Meanwhile, the model based on Dempster-Shafer's theory has figured out Brush Island on the eastern edge of study area (see Figure 3.7), but the island is completely masked by the sea on the map of the Decision Tree, and the Artificial Neural Network again is in between. At the same time, the Artificial Neural Network appears to have over-predicted the Wet *E. maculata* forest and the Dry *E. maculata* forest, and under-predicted the Lower slope wet forest, while the model based on Dempster-Shafer's theory appears to have under-predicted the Wet *E. maculata* forest. In addition, the Artificial Neural Network has predicted the power line easement which runs across the study area from south-west to north-east (see Figure 3.7), the model based on Dempster-Shafer's theory is poorer, and the Decision Tree is the poorest.

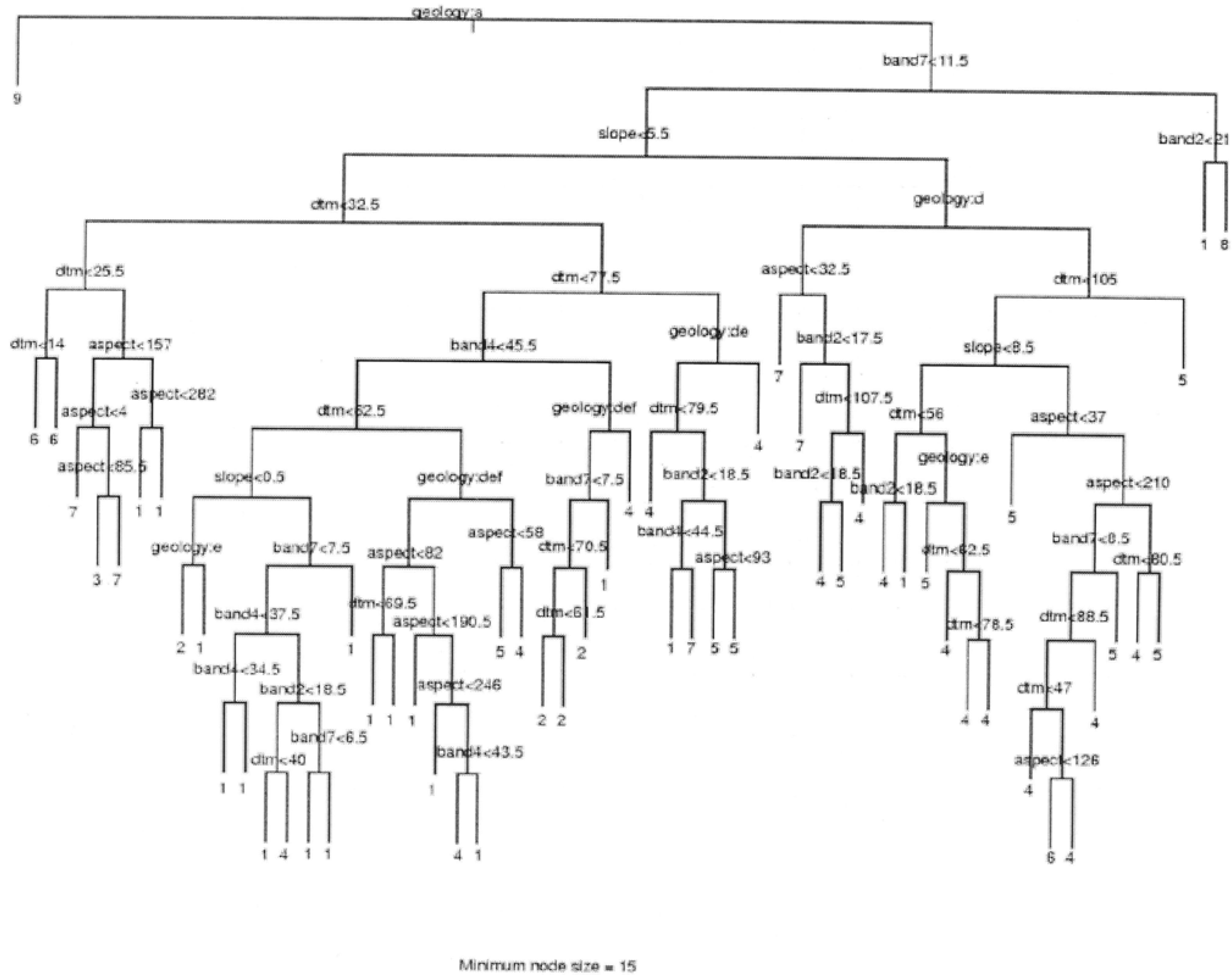


Figure 5.1 The resultant Decision Tree Structure

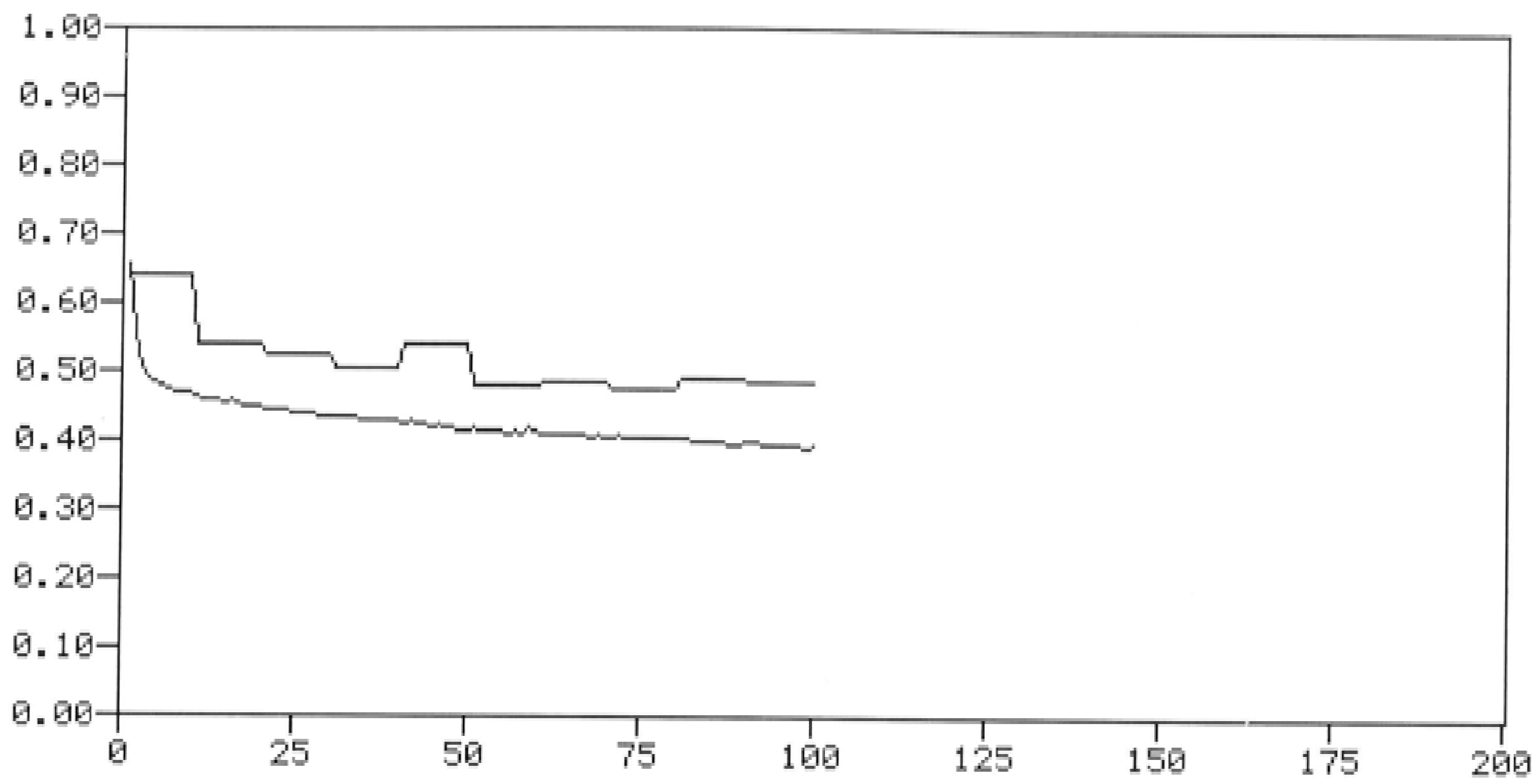


Figure 5.2 The plot of training MSE (lower curve) and testing MSE (upper curve) during the 100 cycles' network training process

Table 5.8 Correlation between the independent variables

Corrlation ®	aspect	band2	Band4	band7	slope	geology	Decision TreeM
aspect	1	-0.022	0.496	0.205	0.357	0.543	0.443
band2		1	0.232	0.568	-0.052	-0.134	-0.1225
band4			1	0.58	0.45	0.692	0.584
band7				1	0.157	0.235	0.133
slope					1	0.388	0.66
geology						1	0.53
DTM							1

Table 5.9 Error matrix of the Decision Tree

		Reference Data										
		1	2	3	4	5	6	7	8	9	Total	User's accuracy
Classified Data	1	**39**	12	6	11	8	3	6	2	0	87	0.45
	2	2	**5**	1	2	1	1	0	0	0	12	0.42
	3	1	0	**2**	0	0	0	0	0	0	3	0.67
	4	11	2	2	**27**	14	8	5	0	0	69	0.39
	5	5	0	1	7	**19**	1	1	1	0	35	0.54
	6	1	0	1	0	1	**6**	0	1	0	10	0.60
	7	1	2	1	0	0	1	**8**	0	0	13	0.62
	8	2	0	1	0	0	0	0	26	0	29	0.90
	9	0	0	0	0	0	0	0	0	89	89	1.0
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.63	0.24	0.13	0.57	0.44	0.30	0.40	0.87	1.0		
	Overall accuracy for 7 forest types	0.464912										
	Kappa accuracy for 7 forest types	0.379422										
	Kappa variance for 7 forest types	0.001406										

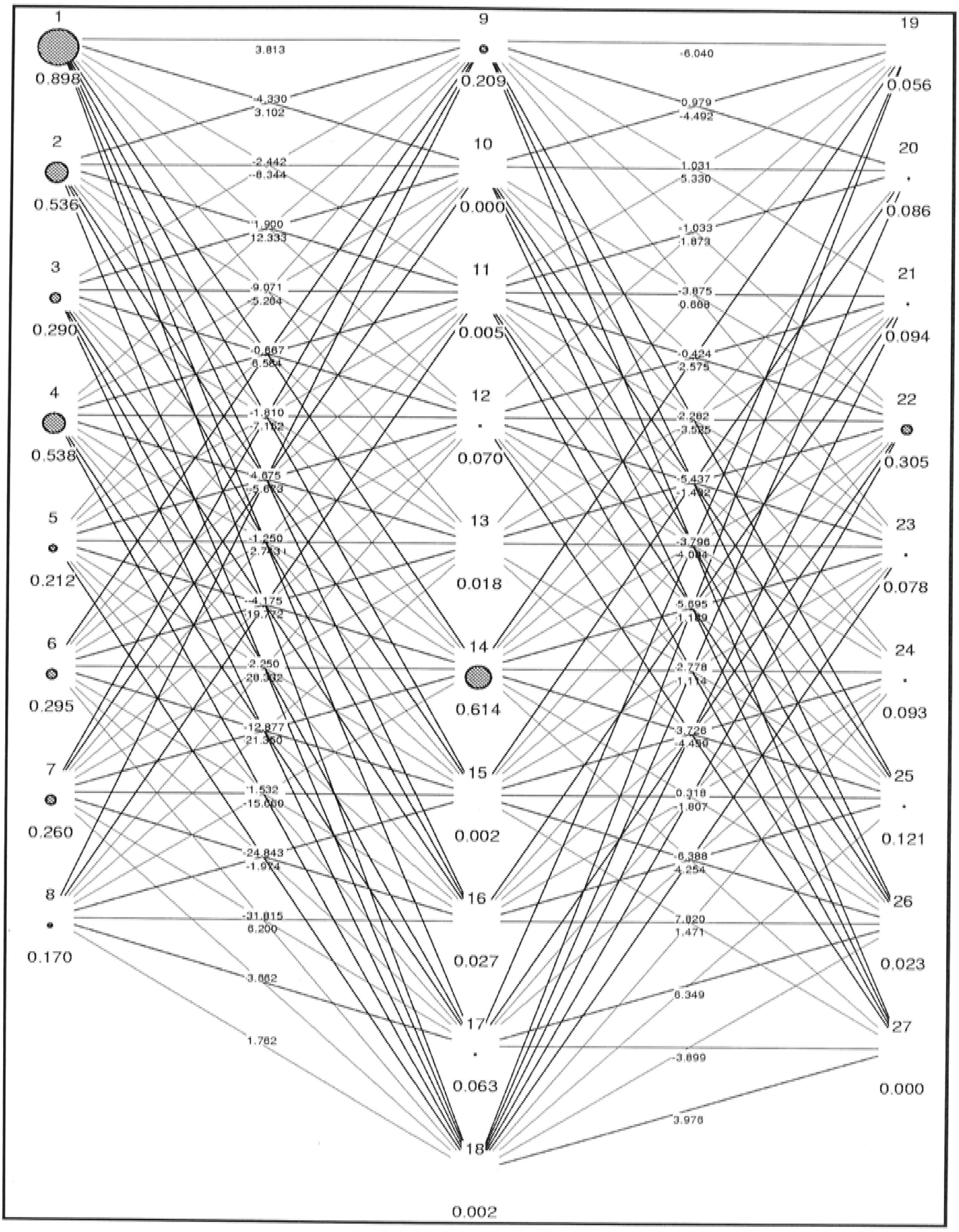


Figure 5.3 The neural network architecture captured at one point of time

Table 5.10 Error Matrix of the Artificial Neural Network

		Reference Data										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**	**User's accuracy**
Classified Data	**1**	**37**	9	6	9	7	1	3	0	0	72	0.51
	2	1	**5**	1	1	1	1	0	0	1	11	0.45
	3	0	0	**1**	0	0	0	0	0	0	1	1.0
	4	15	1	2	**32**	29	11	7	0	0	97	0.33
	5	6	3	1	4	**6**	1	2	0	0	23	0.26
	6	0	1	2	0	0	**6**	1	0	0	10	0.60
	7	1	1	1	1	0	0	**7**	0	0	11	0.64
	8	2	1	1	0	0	0	0	30	1	35	0.86
	9	0	0	0	0	0	0	0	0	87	87	1.0
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.60	0.24	0.07	0.68	0.14	0.30	0.35	1.0	0.98		
	Overall accuracy for 7 forest types	0.412281										
	Kappa accuracy for 7 forest types	0.320582										
	Kappa variance for 7 forest types	0.001285										

Table 5.11 Error matrix of the model based on Dempster-Shafer's theory

		Reference Data										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**	**User's accuracy**
Classified Data	**1**	**29**	4	4	7	4	2	3	1	1	55	0.53
	2	5	**9**	0	1	3	1	1	0	0	20	0.45
	3	11	6	**6**	6	4	4	4	0	0	41	0.15
	4	1	1	0	**16**	8	5	4	0	0	35	0.46
	5	6	1	1	9	**16**	1	0	0	0	34	0.47
	6	8	0	3	6	6	**6**	1	0	0	30	0.20
	7	0	0	0	2	1	0	**7**	0	0	10	0.70
	8	2	0	1	0	1	1	0	29	3	37	0.78
	9	0	0	0	0	0	0	0	0	85	85	1.0
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.47	0.43	0.40	0.34	0.37	0.30	0.35	0.97	0.96		
	Overall accuracy 7 for forest types	0.390351										
	Kappa accuracy for 7 forest types	0.318396										
	Kappa variance for 7 forest types	0.001256										

5.3 DISCUSSION

The results have indicated that the Decision Tree, the Artificial Neural Network and the model based on Dempster-Shafer's theory were capable classifiers for complicated predictive forest type mapping. But the three individual AI models have predicted the forest types in very different ways. Each of the three models has certain advantages and disadvantages, and none was able to produce a clearly better prediction. The Decision Tree is more understandable than the Artificial Neural Network and the model based on Dempster-Shafer's theory. However, in the Decision Tree, a feature could be tested several times along a path from root to leaf (see Figure 5.1), which imposes questions about its comprehensibility. Due to the hierarchical structure of the Decision Tree, the variable tested on top of the tree has greater effect on the prediction. This is the reason that the Decision Tree could not predict Durras Lake and Brush Island, because the geology variable, which was the first variable to be selected, does not distinguish the two geographic features from the surrounding environment (See Figure 3.4).

The Artificial Neural Network of this study obviously did not achieve optimal status, as both the training MSE and the testing MSE were relatively high (Figure 5.2). However, both the training MSE and the testing MSE have been stable after 100 cycles. Though fine-tuning the network architecture and the network parameters may improve the performance to some extent, this is a time and resource consuming process and may not be cost-effective. Furthermore, the reason that the Artificial Neural Network almost could not predict the Lower slope wet forest is due to the fact that the class has the smallest number of samples and bad sample representation for training. The Artificial Neural Network is very sensitive to limited sample size and inadequate sample representation. This finding has confirmed that of Gopal *et al.* (1999) who stated that there may be a minimum number of training samples required for each class in order to learn effectively.

On the other hand, this study has shown that the model based on Dempster-Shafer's theory was less sensitive to the sample size and the sample representation than the Artificial Neural Network and the Decision Tree. For example, the Artificial Neural Network has achieved only a 7% producer's accuracy for the Lower slope wet forest, followed by the Decision Tree's 13%; when the model based on Dempster-Shafer's theory has managed a 40% accuracy. Based on the results of this study (see Tables 5.9-

11), it was found that the model based on Dempster-Shafer's theory has achieved the most consistent predictive accuracies for all classes, the Artificial Neural Network was the least stable, and the Decision Tree was in between. The reason behind this is that Dempster-Shafer's theory assumes independence of input variables and no one variable can dominate others, which has avoided extreme outcomes. However, the assumption of independence may also be a drawback. Because this assumption of independence was not well satisfied in this study (Table 5.1), this may have contributed to the relatively low overall and Kappa predictive accuracies of the model based on Dempster-Shafer's theory. Because this study applied Dempster-Shafer's theory to classify the nine exhaustive exclusive classes, it did not encounter the limitation of computational complexity often tied with Dempster-Shafer's theory when hierarchical evidences are involved. Furthermore, the use of the approximation of Dempster's rule of combination has avoided the problem that may arise when there is zero or one assigned to the mass of support. Generally speaking, given its advantages of simplicity and lesser sensitivity to sampling problems, it is surprising that Dempster-Shafer's theory has not attracted more attention in the classification field.

5.4 SUMMARY

In summary, the author believes that the three individual AI models are moderately good classifiers for complicated forest type mapping with their own advantages and disadvantages. Relying on a single model, and implementing it uncritically, is dangerous. Spending time finding the best model and fine-tuning it is not cost-effective. Perhaps more time should be located on understanding datasets and on studying application.

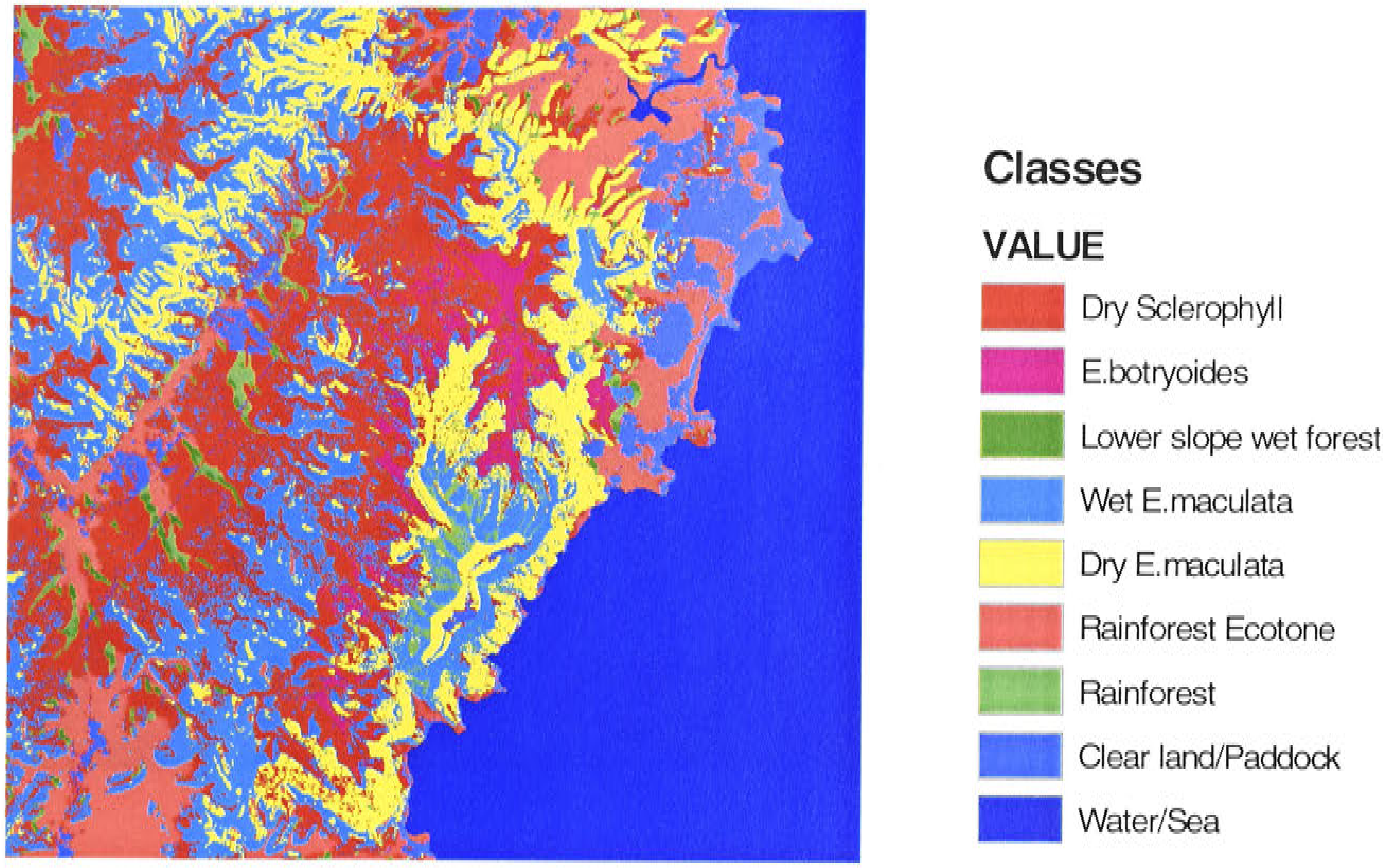


Plate 1: Classification map of the Decision Tree

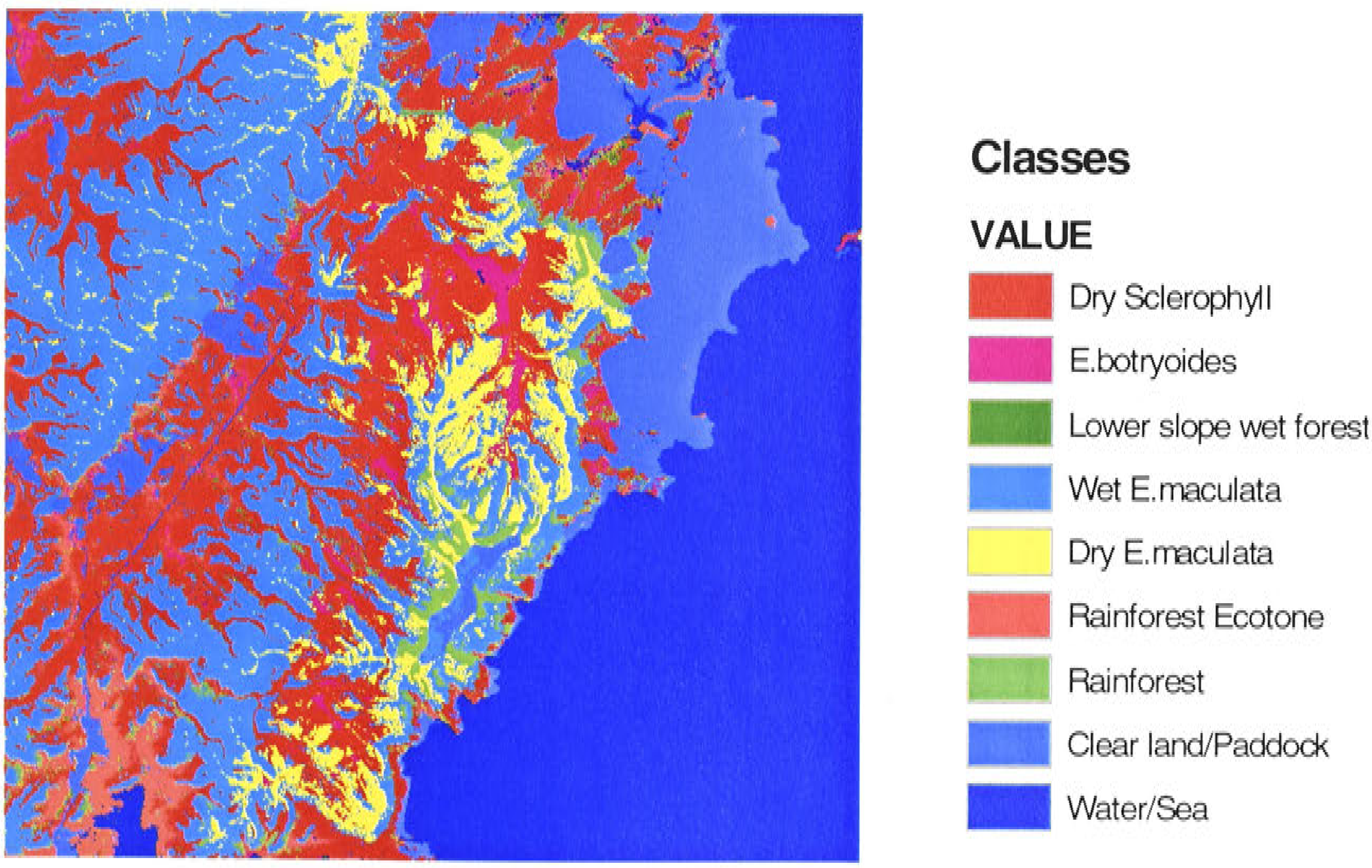


Plate 2: Classification map of the Artificial Neural Network

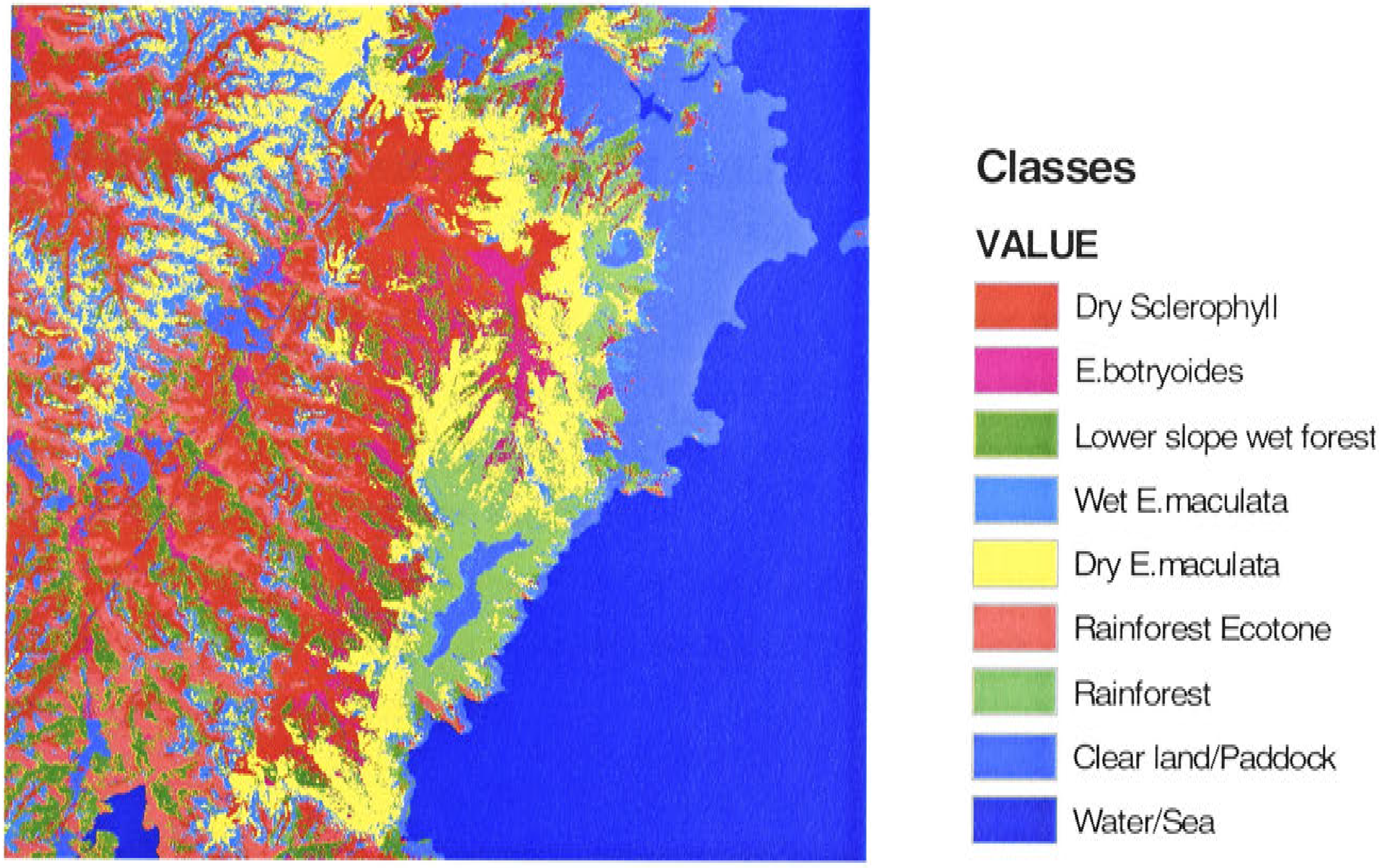


Plate 3: Classification map of the model based on Dempster-Shafer's theory

Chapter 6

COMBINED AI MODELS FOR PREDICTIVE FOREST TYPE MAPPING

The chapter presents the second stage of this study, in which the results of the three individual AI models were combined using different approaches for complicated predictive forest type mapping. Firstly, the chapter describes the combined AI models in detail. Then, the chapter reports the results of the combined AI models for complicated forest type mapping in terms of predictive accuracies and visual appearance. Following that, the chapter discusses the results and findings. Finally, the chapter gives a short summary of this stage of this study.

6.1 METHODS

The three individual AI models applied in the first stage of this study are based on quite different principles. Therefore, they had different characteristics in mapping the complicated forest types as demonstrated in chapter 5. However, by combining the results of these individual AI models, it is possible to improve the classification performance. An assumption for the effectiveness of the combination strategy is that each individual model can provide some unique and useful classification information that is not covered by other models, so by combining them a better model (classifier) can be obtained.

As reviewed in section 2.6, quite a few combination methods have been used and developed in other studies, which range from the simple "production rule" method (Steele, 2000) to a complicated "stack regression analysis" method (Breiman, 1996). This study used different combination approaches to combine the results of the three individual AI models (i.e., the Artificial Neural Network (ANN), the Decision Tree (DT), and the model based on Dempster-Shafer's theory (D-S)) for complicated forest type mapping. The resultant models therefore were called "combined AI models". These combined AI models can be grouped into four subsets dependent on their principles (methods); those based on the majority voting system, the one based on Dempster's rule of combination, those based on simple statistical functions, and those based on fuzzy set theory.

Figure 6.1 displays the flow chart of the combination strategy used in this study. The rectangles represent individual and combined AI models, while the hexagons represent methods used to combine models. The groups of circles represent those combined AI models that are not listed in the diagram due to its limited space or those combined AI models have not been investigated. The solid arrows represent those processes actually implemented in this study, and they are described next. The dashed arrows are some possible applications which have not been investigated in this study.

One important point made by this current study is that combinations can happen at several stages (e.g., this study applied two stages) with potential advantage of further improvement of classification performance at later stages. The other point is that combination approaches can be used to combine more than two individual models. In this study, three individual models were chosen as inputs to the combination approaches. This is because combining only two individual models may not demonstrate the advantages of the combination strategy, while combining more than three individual models may increase the complexity of the combination process.

All combination approaches used in this study are described in detail as follow. It should be noted that they were implemented in the Arc/Info Grid environment.

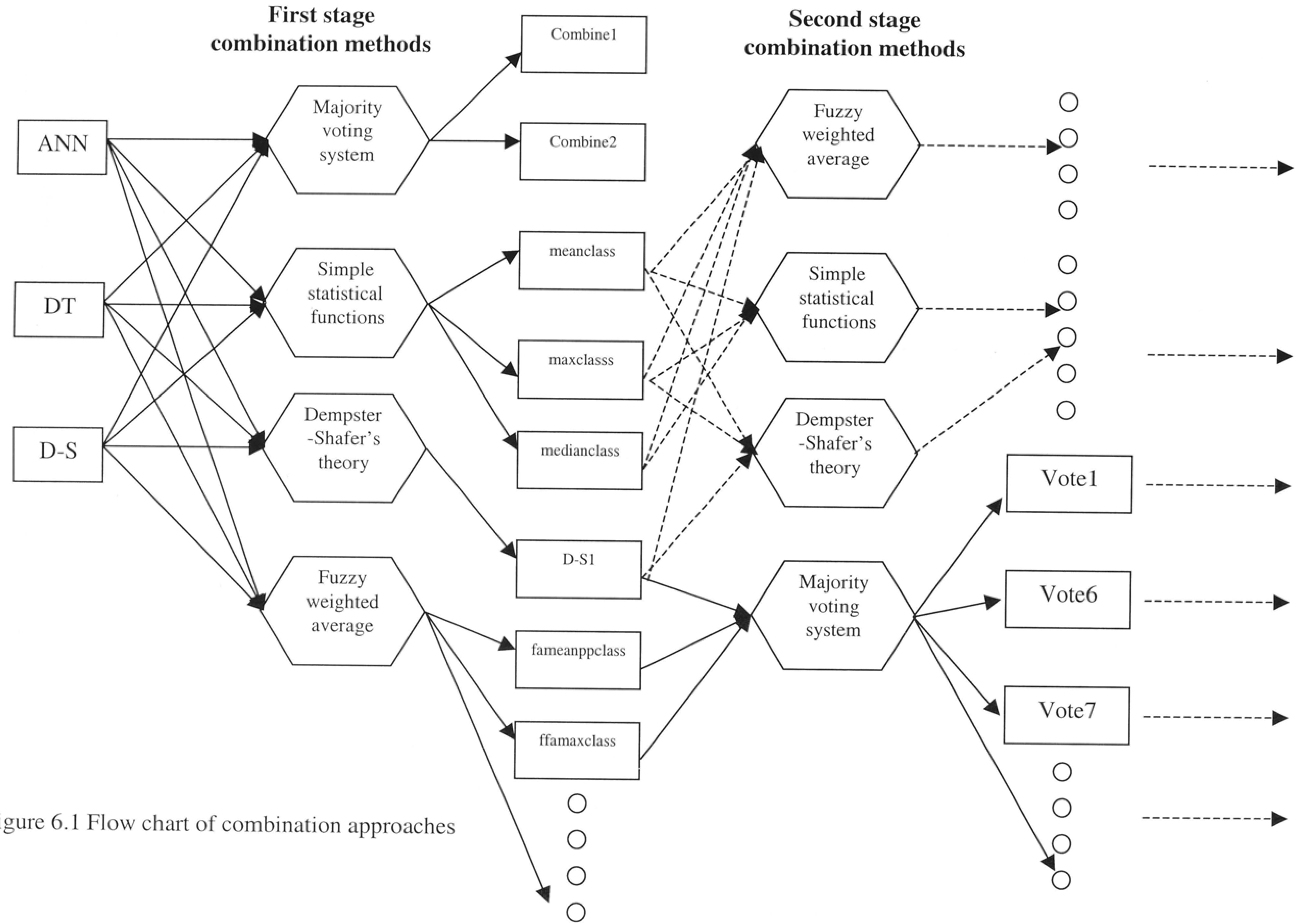


Figure 6.1 Flow chart of combination approaches

6.1.1 The combined AI models based on the majority voting system at the first combination stage

The principle of a majority voting system is often used in decision making processes. The basic idea is that each member of a decision panel has an equal right to vote for a group of options, and the option that obtains the majority of votes wins. However, in the circumstance where none of these options prevail, this rule cannot be applied. Under this situation, sometimes the chair of the decision panel may be given the final say. Sometimes, opinions of certain member of the panel must be obeyed because of his or her expertise. At other times, the decision could be simply suspended due to strong disagreement in the decision panel.

In this study, hard classification results of the Artificial Neural Network, the Decision Tree, and the model based on Dempster-Shafer's theory were combined using two approaches that based on the method of majority voting system (hereafter, they are named combine1 and combine2, see Figure 6.1).

The first approach used the following rules:

If D-S = ANN = DT, then combine1 = D-S;
Else if D-S = ANN $\neq$ DT, then combine1 = D-S;
Else if D-S = DT $\neq$ ANN, then combine1 = D-S;
Else if DT = ANN $\neq$ D-S, then combine1 = DT;
Else if D-S = 2 or D-S = 3, then combine1 = D-S;
Else if ANN = 4 or ANN = 8, then combine1 = ANN;
Else if DT = 1 or DT = 5 or DT = 6 or DT = 7 or DT = 9, then combine1 = DT.

The last rule above indicates a situation that there is no agreement between any panel members. In this case, the model based on Dempster-Shafer's theory was assumed to be 'expert' in predicting class2 and class3, as it obtained the highest producer's accuracies on these two classes; the Artificial Neural Network was assumed to be 'expert' in predicting class4 and class8, as it obtained the highest producer's accuracies on these two classes; and the Decision Tree was assumed to be 'expert' in predicting class1, class5, class6, class7, and class9, as it obtained the highest producer's accuracies on these classes (see Tables 5.9-11).

The second approach used the following rules:

If D-S = ANN = DT, then combine2 = D-S;

Else if D-S = ANN ≠ DT, then combine2 = D-S;

Else if D-S = DT ≠ ANN, then combine2 = D-S;

Else if DT = ANN ≠ D-S, then combine2 = DT;

Else combine2 = DT.

In this case, when there is no agreement between any panel members, the Decision Tree was assumed the chair of the decision panel. It has been given the final say to resolve any disagreement, as it achieved the highest overall predictive accuracies among the three individual AI models (see Tables 5.9-11).

The third option to suspend the decision when there is no agreement among the three classifications was not applied in this study, as the author did not intend to leave any area unclassified.

6.1.2 The combined AI model based on Dempster's rule of combination

If we treat the outcomes of the three AI models as three independent sources of evidence, Dempster's rule of combination can be used to combine them into a single output. In this study, the probability outcomes of the three AI models were combined using the approximation of Dempster's rule of combination described in section 5.1.3. The results were then hardened to produce a conventional classification map (hereafter, the combined AI model is called D-S1, see Figure 6.1).

6.1.3 The combined AI models based on simple statistical functions

Simple statistical functions can be used to combine the probability outcomes of the individual AI models. Three of them used in this study are "MEAN", "MAX", and "MEDIAN". Now, for class 1 at the pixel A of the study area, if we assume that the probability classification outcomes of the three AI models are; P_1(class1|pixelA), P_2(class1|pixelA), and P_3(class1|pixelA). Then, the following three equations based on the three simple statistical functions were used to calculate the outputs of the combined AI models for the class at the pixel:

$$P = MEAN(P_1, P_2, P_3) \tag{6.1}$$

$$P = MAX(P_1, P_2, P_3) \tag{6.2}$$

$$P = MEDIAN(P_1, P_2, P_3) \tag{6.3}$$

The equations were then applied to the whole study area pixel by pixel and class by class. The resultant probability classification outcomes were hardened to produce three conventional classification maps (hereafter, the three combined AI models are named meanclass, maxclass, and medianclass respectively, see Figure 6.1).

6.1.4 The combined AI models based on fuzzy set theory

The above three simple statistical functions are actually special cases of the "weighted average" method. For example, the MAX function assigns full weight to the maximum probability value of three AI models and zero weights to the median and minimum values. The MEDIAN function, on the other hand, assigns full weight to the median value of the three AI models and zero weights to the maximum and minimum values. While, the MEAN function assigns equal one-third weight to each of three models. These simple statistical functions may not be sufficient when considering that the three AI models are based on very different principles, which means that they tend to have different characteristics and predict forest types in very different ways. Therefore, the following approaches have been developed to assign a weight to each of the three AI models based on some **measurements of difference**. All of these combination approaches have utilized the advantages of fuzzy set theory in handling fuzzy terms (fuzzy linguistic variables).

6.1.4.1 The Measurements of Difference

The measurements of difference used in this study can be divided into three groups. The first group looks at the difference between the probability outcome of each model and the mean or median or maximum probability outcome of the three models. It assumes that the mean or median or maximum value is a better estimation of the expected probability outcome. Therefore, the smaller the difference is, the better the agreement is, and the higher fuzzy membership (weight) the model should be assigned. Then a weighted average method can be used to combine the three models, where the fuzzy memberships of individual models play the role of weights. The method is implemented pixel by pixel, class by class. For example, for class 1 at pixel A, we assume that the resultant probability outcomes from the three individual models are P_1(class1|pixel A) = 0.7, P_2(class1|pixelA) = 0.6, and P_3(class1|pixelA) = 0.4. Then under the assumption that the mean value is a better estimation of the expected probability outcome, we obtain:

mean(class1|pixelA) = mean(P_1,P_2,P_3) = 0.57,

difference1 = 0.13,

difference2 = 0.03, and

difference3 = 0.17.

So, model2 should be assigned the highest fuzzy membership (weight) for the pixel and the class, followed by model1 and model3, because the probability outcome of model2 is in a better agreement with the mean value than those of model1 and model3. But under the assumption that the median value is a better estimation, we obtain:

median(class1|pixelA) = median(P_1,P_2,P_3) = 0.6,

difference1 = 0.1,

difference2 = 0, and

difference3 = 0.2.

So, model2 should be again assigned the highest fuzzy membership (weight), but different from the last one, for the pixel and the class, followed by model1 and model3. Under the assumption that the maximum value is a better estimation, we obtain:

maximum(class1|pixelA) = maximum(P_1,P_2,P_3) = 0.7,

difference1 = 0,

difference2 = 0.1, and

difference3 = 0.3.

So, this time model1 should be assigned the highest fuzzy membership (weight) for the pixel and the class, followed by model2 and model3. Hereafter, we call these measurements of difference between the probability outcome of each model and the mean or median or maximum probability outcome of the three models as; bmeanMs, bmedianMs, and bmaxMs, respectively.

The second group of the measurements of difference, however, do not compare the differences between models, but compare the differences within individual models. For example, for any model, for a certain pixel, there are 9 probability outcomes associated with each of the 9 output classes. We assume that the class with the maximum probability outcome is the most likely class for the pixel (i.e., we assign it the highest fuzzy membership (weight) = 1). For each of the other eight classes, the closer its probability outcome to the maximum value, the higher fuzzy membership (weight) it should be assigned. For example, we assume that the 9 probability outcomes for a model at pixel A are; P(class1|pixelA) = P(class2|pixelA) = P(class3|pixelA) = 0.05, P(class4|pixelA) = 0.6, P(class5|pixelA) = 0.2, P(class6|pixelA) = 0.03, P(class7|pixelA) = 0.02, and P(class8|pixelA) = P(class9|pixelA) = 0. Then under the above assumption, we obtain:

maxP(pixelA) = 0.6,

difference1 = difference2 = difference3 = 0.55,

difference4 = 0,

difference5 = 0.3,

difference6 = 0.57,

difference7 = 0.58, and

difference8 = difference9 = 0.6.

So, for this specific model at pixel A, class4 should be assigned the highest fuzzy membership, followed by class5, class1, class2, class3, class6, class7, class8 and class9. The same process can be applied to the other two models throughout the whole study area. Finally, for each class at each pixel, we obtain three fuzzy memberships associated with the three AI models. Then, a weighted average method is applied to combine the three models. Hereafter, we call this measurement of difference within the 9 probability outcomes of a model as; bmaxCs.

The third group of measurements of difference compares the differences among the three models directly, which does not involve calculating the mean, the median and the maximum values. For example, for class 1 at pixel A, we assume that the resultant probability outcomes for the three individual models are P_1(class1|pixel A) = 0.7, P_2(class1|pixelA) = 0.6, and P_3(class1|pixelA) = 0.4. Then we obtain:

difference(P_1,P_2) = 0.1,

difference(P_1,P_3) = 0.3, and

difference(P_2,P_3) = 0.2.

So, the agreement between P_1 and P_2 is the highest and should be assigned the highest fuzzy membership (e.g., $\mu_{agree(P1,P2)}$), followed by the agreement between P_2 and P_3, and the agreement between P_1 and P3. Then the weighted average equation 6.14 (see below) is applied to combine the three models. Hereafter, we call this measurement of difference among models as; bMs.

6.1.4.2 The Fuzzy Membership Functions

To examine if different fuzzy membership functions would make a difference, four groups of fuzzy membership functions have been applied. The first group uses three linear fuzzy membership functions to calculate the agreement level between models. Therefore, a value of measurement of difference can belong to all of the three agreement levels (good agreement, moderate agreement, and low agreement) with different degrees. Figure 6.2 shows the fuzzy membership functions of Equation 6.4, Equation 6.5, and Equation 6.6.

$$\mu_{good-agree} = \begin{cases} 1 - \frac{10 \times x}{3} & x \in [0,0.3) \\ 0 & elsewhere \end{cases} \quad (6.4)$$

$$\mu_{moderate-agree} = \begin{cases} \frac{20 \times x - 2}{3} & x \in [0.1,0.25) \\ \frac{8 - 20 \times x}{3} & x \in [0.25,0.4) \\ 0 & elsewhere \end{cases} \quad (6.5)$$

$$\mu_{low-agree} = \begin{cases} 5 \times x - 1 & x \in [0.2,0.4) \\ 0 & x \in [0,0.2) \\ 1 & elsewhere \end{cases} \quad (6.6)$$

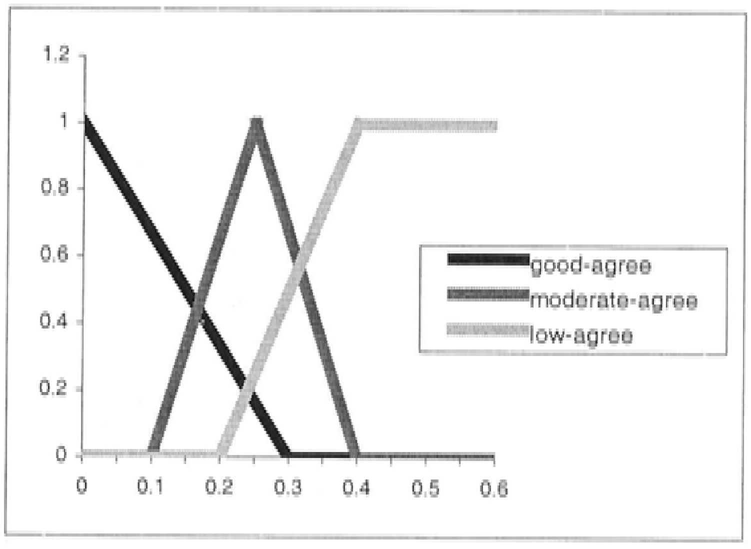


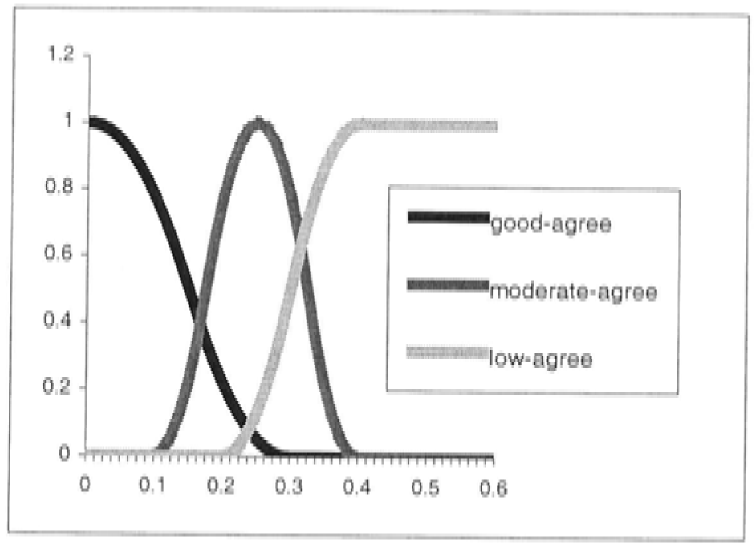
Figure 6.2 Fuzzy membership functions of equation 6.4, equation 6.5, and equation 6.6

where μ is the fuzzy membership, and x is the measurement of difference

The second group instead uses three non-linear fuzzy membership functions to calculate the agreement level between models. Figure 6.3 shows the fuzzy membership functions of Equation 6.7, Equation 6.8, and Equation 6.9.

$$\mu_{good-agree} = \begin{cases} 1 - 2 \times (\frac{x}{0.3})^2 & x \in [0,0.15) \\ 2 \times (\frac{0.3 - x}{0.3})^2 & x \in [0.15,0.3) \\ 0 & elsewhere \end{cases} \quad (6.7)$$

$$\mu_{moderate-agree} = \begin{cases} 2\times(\frac{x-0.1}{0.15})^2 & x\in[0.1,0.175) \\ 1-2\times(\frac{0.25-x}{0.15})^2 & x\in[0.175,0.25) \\ 1-2\times(\frac{x-0.25}{0.15})^2 & x\in[0.25,0.325) \\ 2\times(\frac{0.4-x}{0.15})^2 & x\in[0.325,0.4) \\ 0 & elsewhere \end{cases} \quad (6.8)$$



$$\mu_{low-agree} = \begin{cases} 2\times(\frac{x-0.2}{0.2})^2 & x\in[0.2,0.3) \\ 1-2\times(\frac{0.4-x}{0.2})^2 & x\in[0.3,0.4) \\ 0 & x\in[0,0.2) \\ 1 & elsewhere \end{cases} \quad (6.9)$$

Figure 6.3 Fuzzy membership functions of equation 6.7, equation 6.8, and equation 6.9

where μ is the fuzzy membership, and x is the measurement of difference

The third group uses only one fuzzy membership function to calculate the agreement level between models (Figure 6.4, Equation 6.10):

$$\mu_{agree} = \begin{cases} \frac{0.4-x}{0.4} & x\in[0,0.4) \\ 0 & elsewhere \end{cases} \quad (6.10)$$

where μ is the fuzzy membership, and x is the measurement of difference

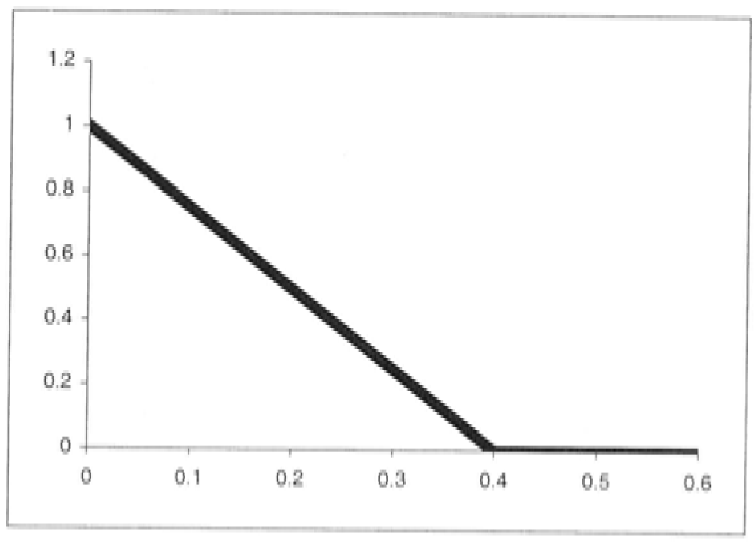


Figure 6.4 Fuzzy membership function of equation 6.10

The fourth group also uses only one fuzzy membership function to calculate the agreement level between models, but this time it is a non-linear function (Figure 6.5, Equation 6.11):

$$\mu_{agree} = \begin{cases} 1-2\times(\frac{x}{0.4})^2 & x\in[0,0.2) \\ 2\times(\frac{0.4-x}{0.4})^2 & x\in[0.2,0.4) \\ 0 & elsewhere \end{cases} \quad (6.11)$$

where μ is the fuzzy membership, and x is the measurement of difference

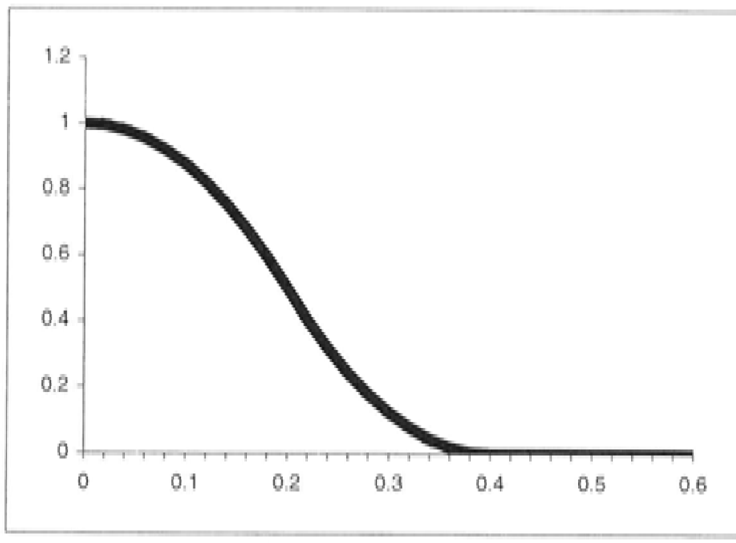


Figure 6.5 Fuzzy membership function of equation 6.11

6.1.4.3 The Weighted Average Equations

So far, we have described the five measurements of difference and the four groups of fuzzy membership functions. Together, eighteen combined AI models based on fuzzy set theory have been developed, which are shown in Table 6.1.

Table 6.1 The eighteen combined AI models based on fuzzy set theory

	bmeanMs	**bmedianMs**	**bmaxMs**	**bmaxCs**	**bMs**
First group	Fmeanclass[1]	Fmedianclass[1]	Fmaxclass[1]	Fcmaxclass[1]	N/A
Second group	Fmeanppclass[1]	Fmedppclass[1]	Fmaxppclass[1]	Fffmaxclass[1]	N/A
Third group	Fameanclass[2]	Famedianclass[2]	Famaxclass[2]	Fffamaxclass[2]	Ffagreeclass[3]
Fourth group	Fameanppclass[2]	Famedppclass[2]	Famaxpclass[2]	Ffamaxclass[2]	Ffappclass[3]

Note that superscript 1 indicates that the model uses the weighted average equation 6.12, superscript 2 indicates that the model uses the weighted average equation 6.13, and superscript 3 indicates that the model uses the weighted average equation 6.14.

After calculating the fuzzy membership values of the three individual AI models, their probability outcomes were combined using three weighted average equations pixel by pixel and class by class. Equation 6.12 was used when involving the first and the second group of fuzzy membership functions, while equation 6.13 was used when involving the third and the fourth group of fuzzy membership functions except when the measurement of difference is bMs, which then used equation 6.14 (see Table 6.1).

These weighted average equations are:

$$P=\begin{cases} \dfrac{P_1+P_2+P_3}{3} & \mu_{good\text{-}agree}=\mu_{moderate\text{-}agree}=\mu_{low\text{-}agree}=0 \\ \dfrac{\sum_{i=1}^{3}(\mu_{good\text{-}agree}(x_i)\times 1+\mu_{moderate\text{-}agree}(x_i)\times 0.5+\mu_{low\text{-}agree}(x_i)\times 0.1)\times P_i}{\sum_{i=1}^{3}(\mu_{good\text{-}agree}(x_i)\times 1+\mu_{moderate\text{-}agree}(x_i)\times 0.5+\mu_{low\text{-}agree}(x_i)\times 0.1)} & elsewhere \end{cases} \tag{6.12}$$

where μ is the fuzzy membership, and x_i are the values of the measurement of difference, P_i are the probability outcomes of the three models, P is the resultant output from the combined models, and 1, 0.5 and 0.1 are the weights assigned to the agreement levels of "good agreement", "moderate agreement", and "low agreement" by the author;

$$P=\begin{cases} \dfrac{P_1+P_2+P_3}{3} & \mu_{agree}=0 \\ \dfrac{\sum_{i=1}^{3}\mu_{agree}(x_i)\times P_i}{\sum_{i=1}^{3}\mu_{agree}(x_i)} & \mu_{agree}\neq 0 \end{cases} \tag{6.13}$$

where μ is the fuzzy membership, and x_i are the values of the measurement of difference, P_i are the probability outcomes of the three models, and P is the resultant output from combined models; and

$$P=\begin{cases} \dfrac{P_1+P_2+P_3}{3} & \mu_{agree}=0 \\ \dfrac{(P_1+P_2)\times\mu_{agree}(P_1,P_2)+(P_1+P_3)\times\mu_{agree}(P_1,P_3)+(P_2+P_3)\times\mu_{agree}(P_2,P_3)}{2\times(\mu_{agree}(P_1,P_2)+\mu_{agree}(P_1,P_3)+\mu_{agree}(P_2,P_3))} & \mu_{agree}\neq 0 \end{cases} \tag{6.14}$$

where μ is the fuzzy membership, and x_i are the values of the measurement of difference, P_i are the probability outcomes of the three models, and P is the resultant output from combined models.

After the probability outcomes of these combined AI models were calculated they were hardened to produce conventional classification maps.

6.1.5 The combined AI models based on the majority voting system at the second combination stage

The results of the above-mentioned combined AI models can be again combined at the second stage by using similar methods. However, only the method of majority voting system was examined in this study, which requires inputs of hard classifications. Three such hardened classifications were chosen, and they are D-S1, fameanppclass, and ffamaxclass (see Figure 6.1).

The first second stage combination approach based on the majority voting system used the following rules:

If D-S1 = fameanppclass = ffamaxclass, then vote1 = D-S1;

Else if D-S1 = 1 or D-S1 = 3 or D-S1 = 5 or D-S1 = 8 or D-S1 = 9, then vote1 = D-S1;

Else if ffamaxclass = 6 or ffamaxclass = 7, then vote1 = ffamaxclass;

Else if fameanppclass = 2 or fameanppclass = 4, then vote1 = fameanppclass;

Else vote1 = D-S1.

The second approach used the following rules:

If D-S1 = fameanppclass = ffamaxclass, then vote6 = D-S1;

Else if D-S1 = 1 or D-S1 = 3 or D-S1 = 5 or D-S1 = 9, then vote6 = D-S1;

Else if ffamaxclass = 6 or ffamaxclass = 7, then vote6 = ffamaxclass;

Else if fameanppclass = 2 or fameanppclass = 4, then vote6 = fameanppclass;

Else vote6 = ffamaxclass.

The third approach used the following rules:

If D-S1 = fameanppclass = ffamaxclass, then vote7 = D-S1;

Else if D-S1 = 1 or D-S1 = 3 or D-S1 = 5 or D-S1 = 9, then vote7 = D-S1;

Else if ffamaxclass = 6 or ffamaxclass = 7 or ffamaxclass = 4, then vote7 = ffamaxclass;

Else if fameanppclass = 2, then vote7 = fameanppclass;

Else vote7 = ffamaxclass.

The three combination approaches all have taken into account the resultant producer's accuracies of D-S1, ffamaxclass and fameanppclass (see Tables 6.4, 6.9, 6.10). They have also considered the different characteristics of the three input models in handling the two known data errors (see section 7.2).

6.2 RESULTS

The outcomes of the three individual AI models were combined in an attempt to improve the classification performance. This section reports the results of the combined AI models for predictive forest type mapping using methods such as the majority voting system, Dempster's rule of combination, simple statistical functions, and fuzzy set theory.

6.2.1 Results of combined AI models based on the majority voting system at the first combination stage

At the first combination stage, two combined AI models based on the majority voting system were applied to combine the hardened classification results of the Decision Tree, the Artificial Neural Network, and the model based on Dempster-Shafer's theory. The error matrix of combine1 is shown in Table 6.2, while the error matrix of combine2 is shown in Table 6.3. It can be seen that combine1 obtained slightly lower overall and Kappa accuracies than those of the Decision Tree, but combine2 achieved an overall accuracy of almost 3.1% and an Kappa accuracy of over 3.6% more than those of the Decision Tree. Moreover, comparing combine2 and the Decision Tree in terms of the user's accuracy and the producer's accuracy, combine2 has obtained higher or equal accuracies for six out of seven forest types except for forest type 5 (Dry *E. maculata*).

Plate 4 and Plate 5 display the classification maps of combine1 and combine2 respectively. They appear to be very similar, but they look quite different from the three initial classifications. Many good features of the three classifications have been retained. Both classifications have nicely predicted Durras Lake and the power line easement and partly predicted Brush Island and Willinga River. In addition, both classifications seem to have neither over-predicted nor under-predicted the Lower slope wet forest and the Wet *E. maculata* forest.

Table 6.2 Error matrix of combine1

		Reference Data										
		1	2	3	4	5	6	7	8	9	Total	User's accuracy
Classified Data	1	**36**	8	4	9	5	3	4	1	0	70	0.51
	2	2	**7**	0	2	2	1	1	0	0	15	0.47
	3	9	4	**4**	0	3	0	2	0	0	22	0.19
	4	9	1	3	**30**	19	9	5	0	0	76	0.39
	5	3	1	1	5	**14**	1	1	0	0	26	0.54
	6	1	0	2	0	0	**6**	0	0	0	9	0.67
	7	0	0	0	1	0	0	**7**	0	0	8	0.88
	8	2	0	1	0	0	0	0	29	1	33	0.88
	9	0	0	0	0	0	0	0	0	88	88	1.0
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.58	0.33	0.27	0.64	0.33	0.30	0.35	0.97	0.99		
	Overall accuracy for 7 forest types	0.45614										
	Kappa accuracy for 7 forest types	0.377317										
	Kappa variance for 7 forest types	0.001355										

Table 6.3 Error matrix of combine2

		Reference Data										
		1	2	3	4	5	6	7	8	9	Total	User's accuracy
Classified Data	1	**44**	10	6	11	10	3	5	1	0	90	0.49
	2	2	**7**	1	1	1	1	0	0	0	13	0.54
	3	1	0	**2**	0	0	0	0	0	0	3	0.67
	4	6	2	2	**29**	15	9	5	0	0	68	0.43
	5	5	1	1	5	**16**	1	1	0	0	30	0.53
	6	1	0	2	0	1	**6**	0	0	0	10	0.60
	7	1	1	0	1	0	0	**9**	0	0	12	0.75
	8	2	0	1	0	0	0	0	29	1	33	0.88
	9	0	0	0	0	0	0	0	0	88	88	1.0
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.71	0.33	0.13	0.62	0.37	0.30	0.45	0.97	0.99		
	Overall accuracy for 7 forest types	0.495614										
	Kappa accuracy for 7 forest types	0.415671										
	Kappa variance for 7 forest types	0.001393										

6.2.2 Results of the combined AI model based on Dempster's rule of combination

Dempster's rule of combination was used to combine the probability outcomes of the three individual models. Table 6.4 shows the error matrix of D-S1, which reports an over 4.8% increase of overall accuracy and an almost 5.9% increase of Kappa accuracy over the Decision Tree. Given the difficulty of predicting these easily confused forest types, the results are impressive. The classification map of D-S1 is shown in Plate 6. It indicates that the combined AI model of D-S1 has nicely predicted Durras Lake, Brush Island, and the power line easement, and it has also partly predicted Willinga River. This represents further improvement of classification performance over the combined AI model of combine2.

Table 6.4 Error matrix of D-S1

		Reference Data										
		1	2	3	4	5	6	7	8	9	Total	User's accuracy
Classified Data	1	**44**	9	8	12	6	3	5	0	1	88	0.50
	2	3	**8**	1	2	1	1	1	0	0	17	0.47
	3	1	3	**3**	0	0	0	0	0	0	7	0.43
	4	2	0	0	**27**	13	9	5	0	0	56	0.48
	5	8	1	1	5	**21**	1	0	0	0	37	0.57
	6	2	0	1	0	1	**5**	0	0	0	9	0.56
	7	0	0	0	1	0	1	**9**	0	0	11	0.82
	8	2	0	1	0	1	0	0	30	3	37	0.81
	9	0	0	0	0	0	0	0	0	85	85	1.0
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.71	0.38	0.20	0.57	0.49	0.25	0.45	1.0	0.96		
	Overall accuracy for 7 forest types	0.513158										
	Kappa accuracy for 7 forest types	0.438323										
	Kappa variance for 7 forest types	0.001394										

6.2.3 Results of the combined AI models based on simple statistical functions

Three simple statistical functions of "MEAN", "MAX", and "MEDIAN" were used to combine the probability outcomes of the three individual AI models. The error matrix of meanclass is shown in Table 6.5, which reports an over 3.5% increase of overall accuracy and over 4.4% increase of Kappa accuracy from those of the Decision Tree. The error matrix of maxclass is shown in Table 6.6, which reports a nearly 1.8% increase of overall accuracy and a nearly 2.6% increase of Kappa accuracy from those of the Decision Tree. The error matrix of medianclass is shown in Table 6.7, which reports a nearly 2.2% increase of overall accuracy and a nearly 3.0% increase of Kappa accuracy from those of the Decision Tree. These results indicate that all of the three combined AI models based on simple statistical functions were able to increase predictive accuracies. Among them the combined AI model using the simple MEAN function is the best. Meanclass has also achieved slightly higher overall and Kappa accuracies than combine2 of the majority voting system, but they are not as good as those of D-S1 of Dempster's rule of combination.

Table 6.5Error matrix of meanclass

Classified Data		Reference Data										
		1	2	3	4	5	6	7	8	9	Total	User's accuracy
	1	**43**	10	8	12	7	4	4	1	0	89	0.48
	2	3	**8**	1	2	1	1	1	0	0	17	0.47
	3	1	2	**3**	0	1	0	1	0	0	8	0.38
	4	2	0	0	**27**	13	8	5	0	0	55	0.49
	5	8	1	1	5	**19**	1	0	0	0	35	0.54
	6	3	0	1	0	1	**5**	0	0	0	10	0.50
	7	0	0	0	1	0	1	**9**	0	0	11	0.82
	8	2	0	1	0	1	0	0	29	1	34	0.85
	9	0	0	0	0	0	0	0	0	88	88	1.00
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.69	0.38	0.20	0.57	0.44	0.25	0.45	0.97	0.99		
	Overall accuracy for 7 forest types	0.50										
	Kappa accuracy for 7 forest types	0.423445										
	Kappa variance for 7 forest types	0.001393										

Table 6.6 Error matrix of maxclass

		Reference Data										
		1	2	3	4	5	6	7	8	9	Total	User's accuracy
Classified Data	1	**40**	12	7	12	5	3	4	1	0	84	0.48
	2	1	**5**	0	2	1	1	0	0	0	10	0.50
	3	2	1	**3**	0	3	0	1	0	0	10	0.30
	4	6	0	1	**25**	12	6	4	0	0	54	0.46
	5	7	1	1	6	**20**	1	1	0	0	37	0.54
	6	3	0	1	1	1	**7**	0	0	0	13	0.54
	7	1	2	1	1	0	1	**10**	0	0	16	0.63
	8	2	0	1	0	1	1	0	29	2	36	0.81
	9	0	0	0	0	0	0	0	0	87	87	1.00
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.65	0.24	0.20	0.53	0.47	0.35	0.50	0.97	0.98		
	Overall accuracy for 7 forest types	0.482456										
	Kappa accuracy for 7 forest types	0.405192										
	Kappa variance for 7 forest types	0.001392										

Table 6.7 Error matrix of medianclass

		Reference Data										
		1	2	3	4	5	6	7	8	9	Total	User's accuracy
Classified Data	1	**42**	8	6	11	6	2	4	1	0	80	0.53
	2	3	**9**	2	1	1	1	1	0	0	18	0.50
	3	1	1	**2**	0	0	0	0	0	0	4	0.50
	4	4	2	0	**28**	18	9	6	0	0	67	0.42
	5	8	1	1	6	**16**	1	1	0	0	34	0.47
	6	2	0	2	0	1	**6**	0	0	0	11	0.55
	7	0	0	1	1	0	0	**8**	0	0	10	0.80
	8	2	0	1	0	1	1	0	29	1	35	0.83
	9	0	0	0	0	0	0	0	0	88	88	1.00
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.68	0.43	0.13	0.60	0.37	0.30	0.40	0.97	0.99		
	Overall accuracy for 7 forest types	0.486842										
	Kappa accuracy for 7 forest types	0.408927										
	Kappa variance for 7 forest types	0.001385										

The classification maps of the three combined AI models are shown in Plates 7-9 respectively. It appears that the classification maps of meanclass and medianclass are very similar. They both have predicted Durras Lake and the power line easement very well, and they have predicted Brush Island and Willinga River partly. On the other hand, the classification map of maxclass looks a bit different from the above two, which predicted Brush Island slightly better than the above two but still not as well as D-S1.

6.2.4 Results of the combined AI models based on fuzzy set theory

The combined AI models based on fuzzy set theory used weighted-average method to combine the probability outcomes of the three individual models. Consequently, eighteen such combined AI models were developed on the base of five **measurements of difference** and four groups of fuzzy membership functions. The overall accuracies and Kappa accuracies of these combined AI models are summarized in Table 6.8.

Table 6.8 The overall accuracies and Kappa accuracies of the combined AI models based on fuzzy set theory

	Name	***Measurements of difference***				
	Overall accuracy / **Kappa accuracy**	**bmeanMs**	**bmedianMs**	**bmaxMs**	**bmaxCs**	**BMs**
Fuzzy membership functions	**First group**	Fmeanclass	Fmedianclass	Fmaxclass	Fcmaxclass	N/A
		0.482456	0.473684	0.486842	0.504386	
		0.403962	0.394468	0.410592	0.428436	
	Second group	Fmeanppclass	Fmedppclass	Fmaxppclass	Fffmaxclass	N/A
		0.482456	0.47807	0.482456	0.50	
		0.40338	0.399916	0.405813	0.424912	
	Third group	Fameanclass	Famedianclass	Famaxclass	Fffamaxclass	Ffagreeclass
		0.47867	0.486842	0.482456	0.508772	0.477807
		0.399296	0.409538	0.405813	0.43496	0.399139
	Fourth group	Fameanppclass	Famedppclass	Famaxpclass	Ffamaxclass	Ffappclass
		0.495614	0.486842	0.491228	0.513158	0.491228
		0.41964	0.408186	0.415689	0.440241	0.413518

It shows that all these combined AI models based on fuzzy set theory have increased the predictive accuracies from those of the Decision Tree to some extent. These results also indicate that among the five measurements of difference, the combined AI models based on bmaxCs or based on the measurement of difference within the nine probability outcomes of a model achieved consistently higher predictive accuracies than those combined AI models based on the other four measurements of difference. Moreover, among the four groups of fuzzy membership functions being used, the combined AI models utilising the fourth group fuzzy membership functions obtained generally better predictive accuracies than those utilising the other three groups of fuzzy membership functions.

From all these combined AI models based on fuzzy set theory, ffamaxclass and fameanppclass have been chosen for the second stage combination. Their error matrices are shown in Table 6.9 and Table 6.10 respectively.

Table 6.9 Error matrix of ffamaxclass

		Reference Data										
		1	2	3	4	5	6	7	8	9	Total	User's accuracy
Classified Data	1	**43**	11	6	10	7	2	4	1	0	84	0.51
	2	2	**7**	1	3	1	1	0	0	0	15	0.47
	3	1	1	**3**	0	1	0	1	0	0	7	0.43
	4	6	1	2	**27**	11	6	4	0	0	57	0.47
	5	5	1	1	5	**19**	1	1	0	0	33	0.58
	6	2	0	1	1	2	**8**	0	1	0	15	0.53
	7	1	0	0	1	1	1	**10**	0	0	14	0.71
	8	2	0	1	0	1	1	0	28	2	35	0.80
	9	0	0	0	0	0	0	0	0	87	87	1.00
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.69	0.33	0.20	0.57	0.44	0.40	0.50	0.93	0.98		
	Overall accuracy for 7 forest types	0.513158										
	Kappa accuracy for 7 forest types	0.440241										
	Kappa variance for 7 forest types	0.001395										

Table 6.10 Error matrix of fameanppclass

Classified Data		Reference Data 1	2	3	4	5	6	7	8	9	Total	User's accuracy
	1	**41**	8	6	12	5	2	4	1	0	79	0.52
	2	4	**9**	2	2	1	1	1	0	0	20	0.45
	3	1	1	**3**	0	0	0	0	0	0	5	0.60
	4	4	2	0	**28**	17	9	6	0	0	66	0.42
	5	8	1	1	4	**18**	1	1	0	0	34	0.53
	6	2	0	2	0	1	**6**	0	0	0	11	0.55
	7	0	0	0	1	0	0	**8**	0	0	9	0.89
	8	2	0	1	0	1	1	0	29	1	35	0.83
	9	0	0	0	0	0	0	0	0	88	88	1.00
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.66	0.43	0.20	0.60	0.42	0.30	0.40	0.97	0.99		
	Overall accuracy for 7 forest types	0.495614										
	Kappa accuracy for 7 forest types	0.41964										
	Kappa variance for 7 forest types	0.001395										

Table 6.9 reveals that ffamaxclass is the best combined AI model at the first stage combination, which has achieved an over 4.8% higher overall accuracy and a nearly 6.1% higher Kappa accuracy than the Decision Tree. For fameanppclass, the extents of increase are nearly 3.1% on overall accuracy and slightly over 4% on Kappa accuracy respectively.

The classification maps of the eighteen combined AI models are shown in Plates 10-27 respectively. All of them have nicely predicted the power line easement. Among them, the ten combined AI models based on bmeanMs, bmedianMs and bMs have predicted Durras Lake in good shape and predicted Brush Island and Willinga River in part. The rest of eight classification maps based on bmaxMs and bmaxCs look similar to one another. They all have predicted Willinga River nicely and predicted Brush Island in large part. In addition, six of the eight classifications have predicted Durras Lake in large part except for the ffamaxclass and the fffamaxclass which only predicted a small part of the lake.

6.2.5 Results of the combined AI models based on the majority voting system at the second combination stage

The combined AI models of D-S1, ffamaxclass, and fameanppclass from the first stage combination were chosen as inputs to the second stage combination based on the majority voting system. The reason for selecting D-S1 and ffamaxclass is their higher predictive accuracies than the other combined AI models. The reason to select fameanppclass, however, is that it is based on a different principle from that of D-S1 and ffamaxclass. Many other combined AI models are possible but have not been investigated in this study.

Tables 6.11-13 report the error matrices of vote1, vote2, and vote7 respectively. These results clearly show that the three combined AI models at the second stage combination have further increased the predictive accuracies from those of the first stage combination. Among them, vote7 is the best, which has increased an overall accuracy of over 2.6% and a Kappa accuracy of nearly 2.9% from those of ffamaxclass. In other words, the results of vote7 indicate an over 7.4% increase of overall accuracy and a nearly 9% increase of Kappa accuracy over the Decision Tree. These results are significant under the uncertainties and difficulties that exist in predictive forest type mapping. The other achievement of vote7 needs to be mentioned is that six out of seven of its user's accuracies for the 7 forest types are higher or equal to 50%. Its user's accuracy of the Lower slope wet forest is 43% in the case of a very small number of training samples. The significance of this finding means that the classification map of vote7 may serve as a fairly good reference in identifying individual forest types if brought to the field.

Plates 28 - 30 display the classification maps of vote1, vote6 and vote7 respectively. Visually, the classification maps of vote6 and vote7 are very similar. They have perfectly predicted Durras Lake, the power line easement and Willinga River, and they also predicted Brush Island in large part. Vote1 however, has predicted Brush Island nicely, but it only partly predicted Willinga River.

Table 6.11Error matrix of vote1

	Reference Data										
Classified Data	1	2	3	4	5	6	7	8	9	Total	User's accuracy
1	**44**	9	8	12	6	3	5	0	1	88	0.50
2	2	**8**	1	2	1	1	0	0	0	15	0.53
3	1	3	**3**	0	0	0	0	0	0	7	0.43
4	1	0	0	**26**	12	6	5	0	0	50	0.52
5	8	1	1	5	**21**	1	0	0	0	37	0.57
6	3	0	1	1	2	**8**	0	0	0	15	0.53
7	1	0	0	1	0	1	**10**	0	0	13	0.77
8	2	0	1	0	1	0	0	30	3	37	0.81
9	0	0	0	0	0	0	0	0	85	85	1.00
Total	62	21	15	47	43	20	20	30	89	347	
Producer's accuracy	0.71	0.38	0.20	0.55	0.49	0.40	0.50	1.00	0.96		
Overall accuracy for 7 forest types	0.526316										
Kappa accuracy for 7 forest types	0.454807										
Kappa variance for 7 forest types	0.001401										

Table 6.12 Error matrix of vote6

	Reference Data										
Classified Data	1	2	3	4	5	6	7	8	9	Total	User's accuracy
1	**45**	9	8	12	6	3	5	1	1	90	0.50
2	2	**8**	1	2	1	1	0	0	0	15	0.53
3	1	3	**3**	0	0	0	0	0	0	7	0.43
4	1	0	0	**26**	12	6	5	0	0	50	0.52
5	8	1	1	5	**21**	1	0	0	0	37	0.57
6	2	0	1	1	2	**8**	0	1	0	15	0.53
7	1	0	0	1	0	1	**10**	0	0	13	0.77
8	2	0	1	0	1	0	0	28	2	34	0.82
9	0	0	0	0	0	0	0	0	86	86	1.00
Total	62	21	15	47	43	20	20	30	89	347	
Producer's accuracy	0.73	0.38	0.20	0.55	0.49	0.40	0.50	0.93	0.97		
Overall accuracy for 7 forest types	0.530702										
Kappa accuracy for 7 forest types	0.458879										
Kappa variance for 7 forest types	0.001403										

Table 6.13 Error matrix of vote7

Classified Data	Reference Data 1	2	3	4	5	6	7	8	9	Total	User's accuracy
1	**45**	9	8	12	6	3	5	1	1	90	0.50
2	2	**8**	1	2	1	1	0	0	0	15	0.53
3	1	3	**3**	0	0	0	0	0	0	7	0.43
4	1	0	0	**26**	10	6	4	0	0	47	0.55
5	8	1	1	5	**23**	1	1	0	0	40	0.58
6	2	0	1	1	2	**8**	0	1	0	15	0.53
7	1	0	0	1	0	1	**10**	0	0	13	0.77
8	2	0	1	0	1	0	0	28	2	34	0.82
9	0	0	0	0	0	0	0	0	86	86	1.00
Total	62	21	15	47	43	20	20	30	89	347	
Producer's accuracy	0.73	0.38	0.20	0.55	0.53	0.40	0.50	0.93	0.97		
Overall accuracy for 7 forest types	0.539474										
Kappa accuracy for 7 forest types	0.469087										
Kappa variance for 7 forest types	0.001401										

6.3 DISCUSSION

This study has shown that the combination strategy was an effective and efficient strategy for complicated forest type mapping. All of the combined AI models except one have increased the predictive accuracy and improved the visual appearance of the classification maps. The finding has confirmed the hypothesis that each individual model can provide some unique and useful information which is not covered by other models, so by combining them a better model (classifier) can be obtained. This agrees with the findings of Xu *et al.* (1992), Rogova (1994), and See *et al.* (1998). In this study, the combination strategy has acted as a filter with consistent evidence being retained, and conflicting evidence being smoothed.

It can be seen that all of the combined AI models except combine1 have improved the classification performance from the three individual AI models. This has indicated that combination methods based on the majority voting system, simple statistical functions, Dempster's rule of combination, and fuzzy set theory all could be used for combining individual models. It has also demonstrated that the combination could be implemented in several stages, which could further improve the classification performance at the later stages.

It is surprising that combine1 did not increase the predictive accuracies over the Decision Tree as combine2 did, as the two classification maps appear to be so similar. This may again indicate that assessing classifiers relying only on a small size of test samples is not reliable. The visual assessment and the estimation of prediction confidence should be also considered if they are available. It is not unexpected that the classification maps of combine1 and combine2 appear to be very different from those of the three individual AI models, as the combinations tend to have smoothed the differences among the three individual AI models. Though combine2 achieved better predictive accuracies than many of other combined AI models, deferring the hardening process could provide more flexibility. The reason is that hardening the probability outcomes before combining them could cause the loss of useful information.

The results of D-S1 were encouraging. This has been demonstrated not only by the large increases of the predictive accuracies and the consistent user's accuracies (see Table 6.4) but also by the better-predicted classification map (see Plate 6). The visual appearance of D-S1 is among the best of all combined AI models. This has once again confirmed the attractive features of Dempster's rule of combination, in that concordant items of evidence reinforce each other, and conflicting items of evidence erode each other (Shafer, 1990; Murphy, 2000).

The combination approaches of the three simple statistical functions were the most convenient combined AI models to have been implemented. Therefore, they may be more cost-effective than other combined AI models in terms of time and resource requirement, and classification performance. The higher predictive accuracies of meanclass than those of medianclass and maxclass are understandable. Because for each class at each pixel, meanclass has used the inputs of all three individual AI models, while medianclass and maxclass have only used one of the three probability outcomes. This again indicates that each of the three individual AI models has caught some unique and useful information for the classification of the pixel, and rejecting some of the potentially useful information may not be wise.

Given the simplicity of the three simple statistical functions, they may become the start points for those users who want to examine the usefulness of the combination strategy. However, this may not be sufficient when considering that the three individual AI models are based on very different principles, which means that they tend to have different characteristics in predicting forest types. The combination method based on fuzzy set theory was developed to deal with the potential problem. The results have shown that some of these combined AI models based on the fuzzy weighted average approaches did further increase the predictive accuracies. It is not clear why combined AI models based on the measurement of difference of bmaxCs achieved consistently higher predictive accuracies than those based on the other four measurements of difference and those based on the three simple statistical functions. This may well be because of the different principles the three individual AI models are based on. It could indicate that the resultant absolute probability outcomes of the three individual AI models may not be directly comparable. On the other hand, the measures of relative likelihood of one pixel belonging to each class may be better comparison standards for individual models. The better results from the combined AI models using the fourth group fuzzy membership function have shown that non-linear fuzzy membership functions may be better than the linear ones, and the agreement level between models may be better evaluated under one uniform standard (fuzzy membership function) rather than three standards. In addition, it is disappointing that though ffamaxclass has achieved higher predictive accuracies than other combined AI models at the first stage combination, it did not guarantee a better visual appearance, as it almost did not predict Durras Lake.

On the other hand, it is very encouraging to find that both the predictive accuracies and the visual appearance of vote7 have been improved from those of the first stage combination. The significance of an over 7% increase of overall accuracy and a nearly 9% increase of Kappa accuracy over the Decision Tree should not be underestimated. Especially as this improvement was achieved under the uncertainties and difficulties that exist in predictive forest type mapping which, as argued by Lees (1996a), has resulted in an upper limit on predictive accuracy.

6.4 SUMMARY

In summary, the author believes that the combination strategy is effective and efficient in improving the classification performance. Given that many combination approaches are simple and reliable, applying the combination strategy to improve the classification performance under difficult situations is more cost-effective than spending time finding a best single model such as an Artificial Neural Network and fine-tuning the model. However, a pre-requirement for the combination strategy is that there have to be at least three good classifiers available in order for the strategy to be implemented effectively. Fortunately, the scientific field is now developing more and more AI models that can serve as good classifiers, and these models are become increasing commercialized. Combining models based on different principles is better than combining models based on similar principles, because models based on similar principles are likely to have common blind spots for an application, while models based on different principles may compensate for each other.

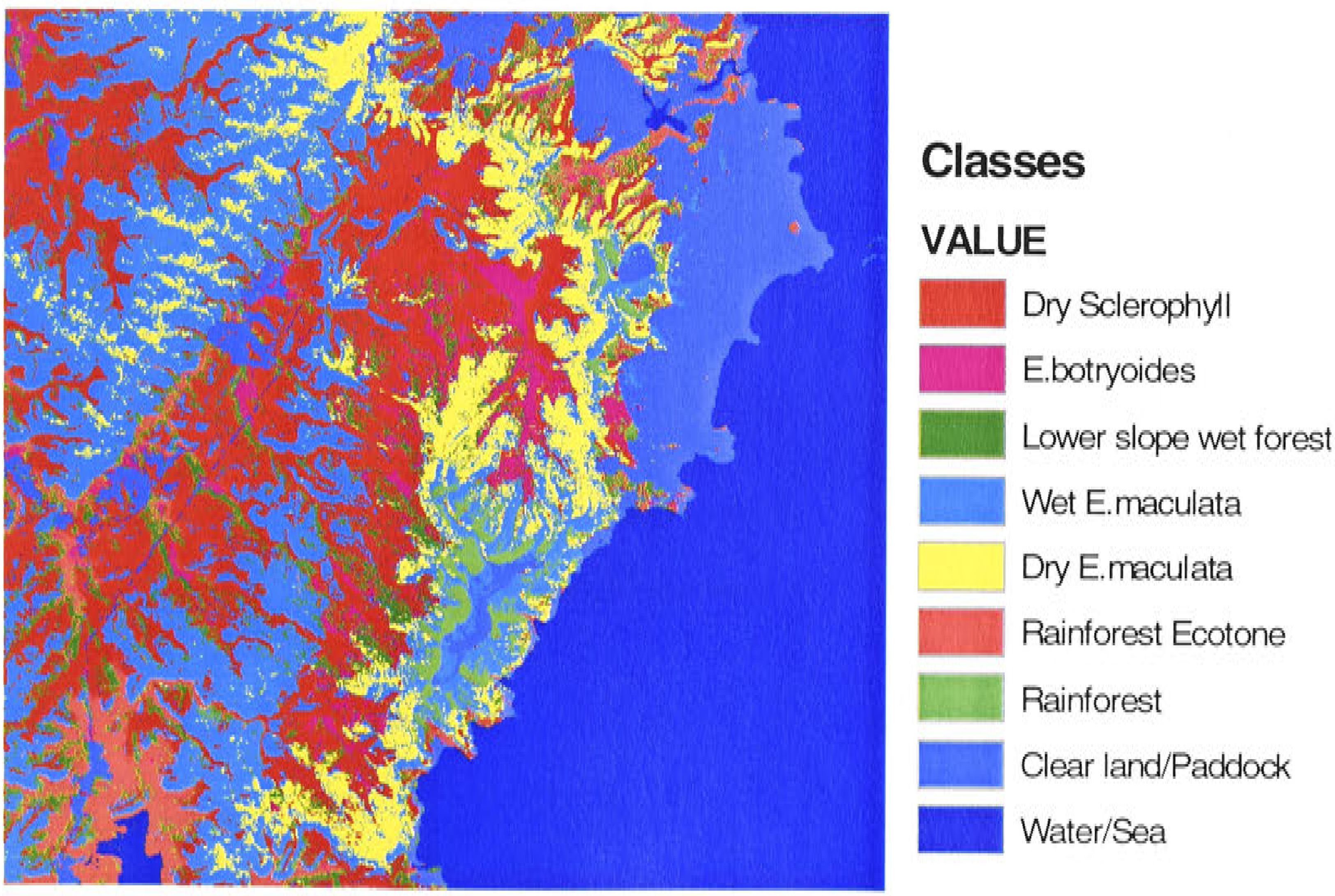


Plate 4 Classification map of combine1

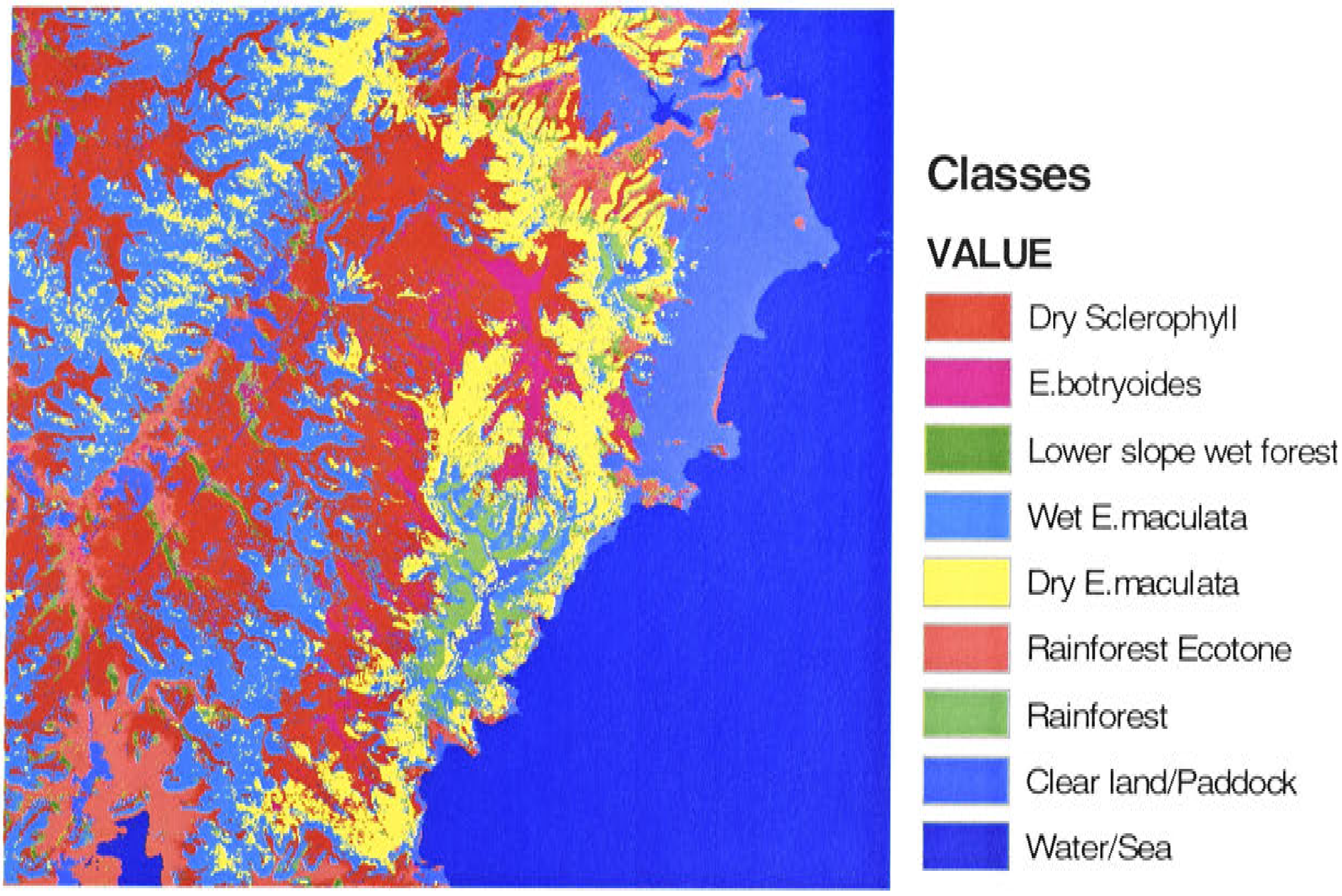


Plate 5 Classification map of combine2

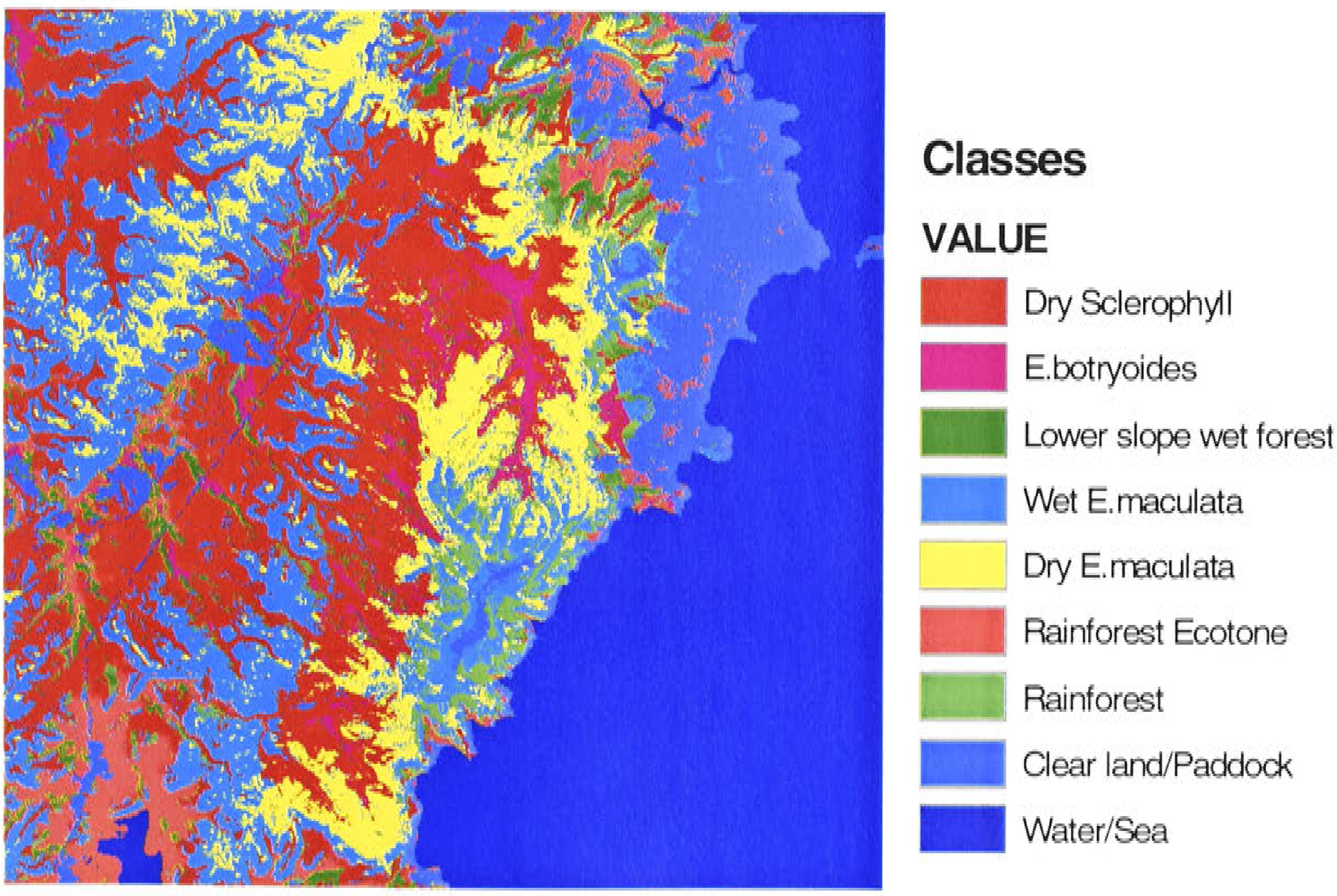


Plate 6 Classification map of D-S1

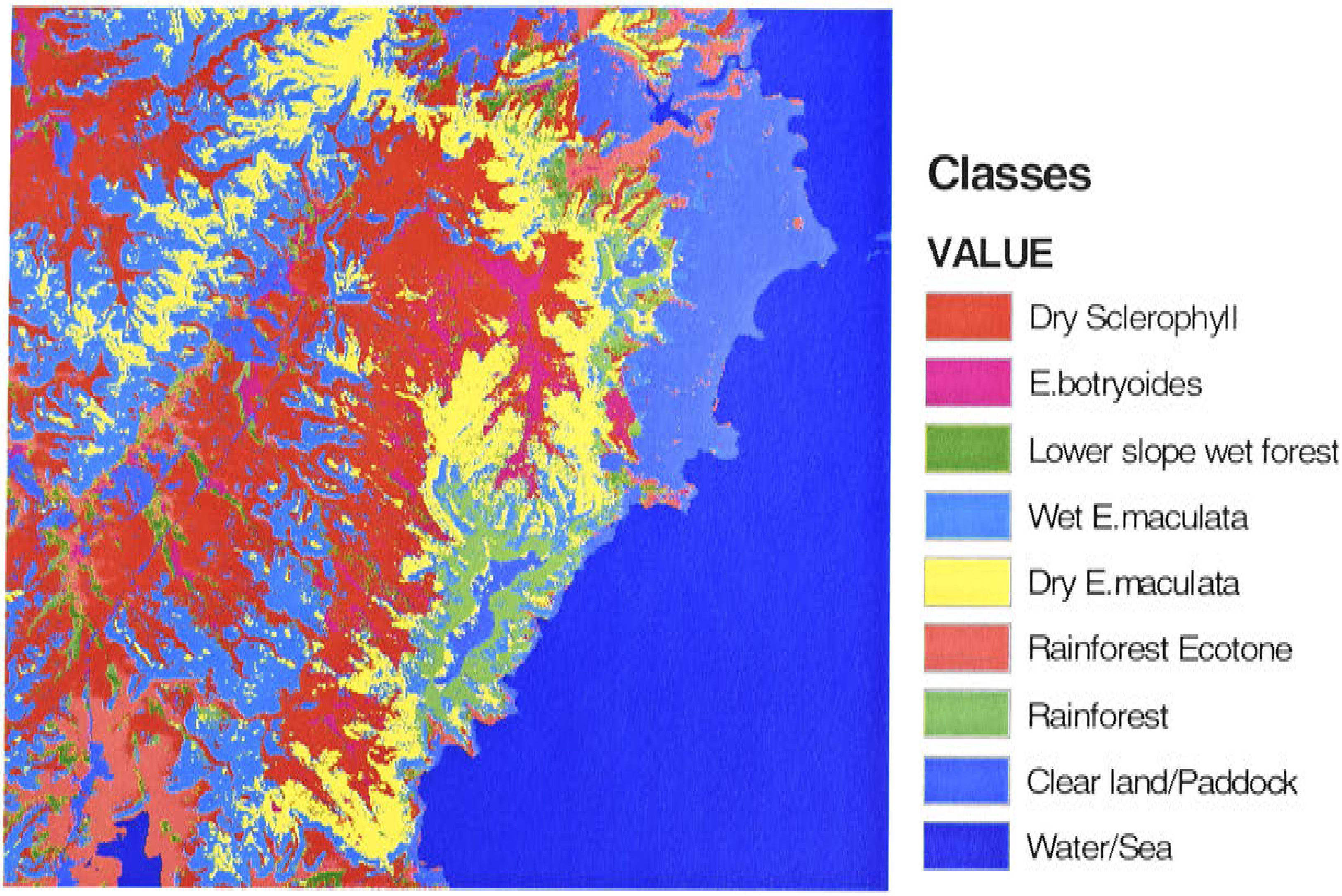


Plate 7 Classification map of meanclass

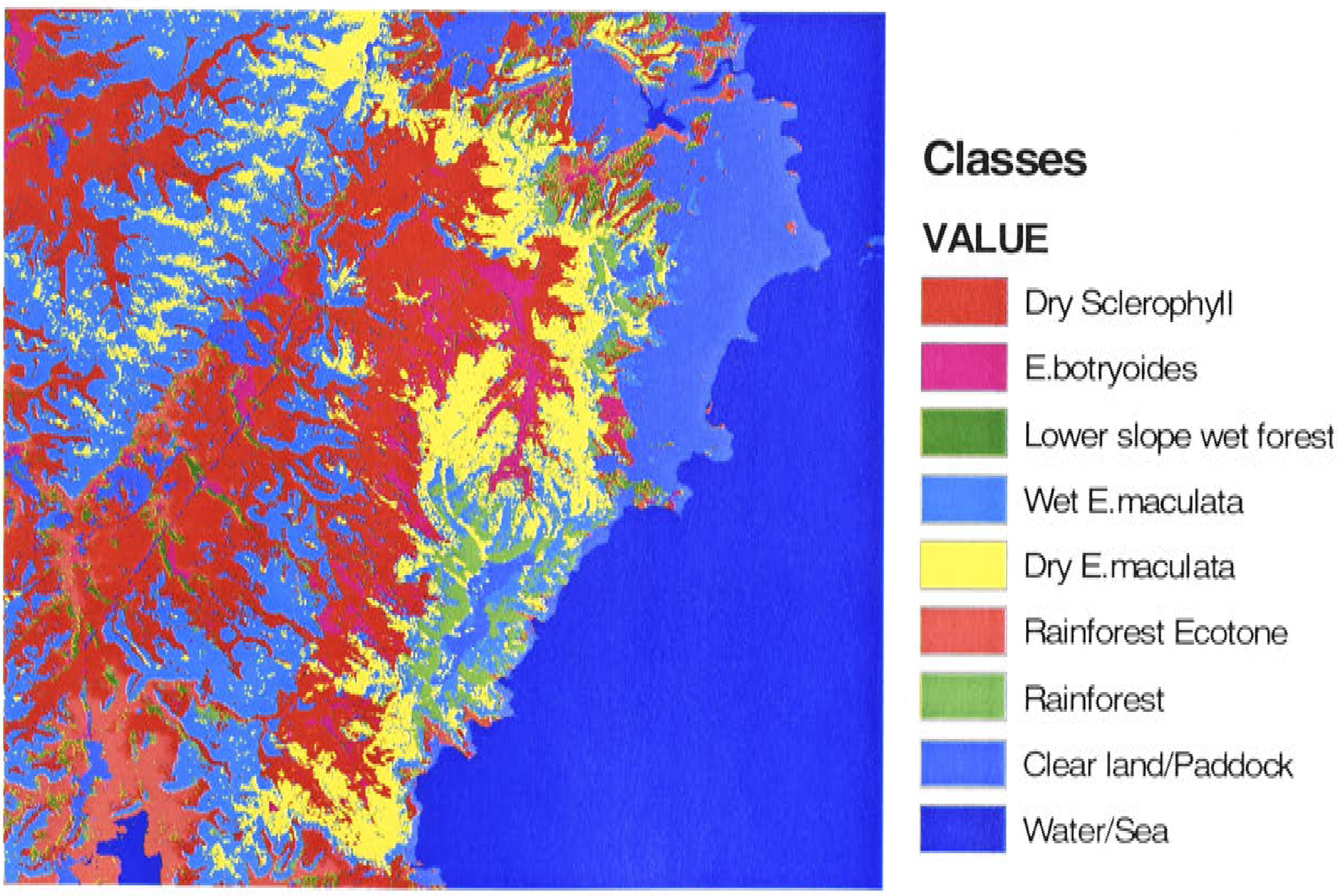


Plate 8 Classification map of medianclass

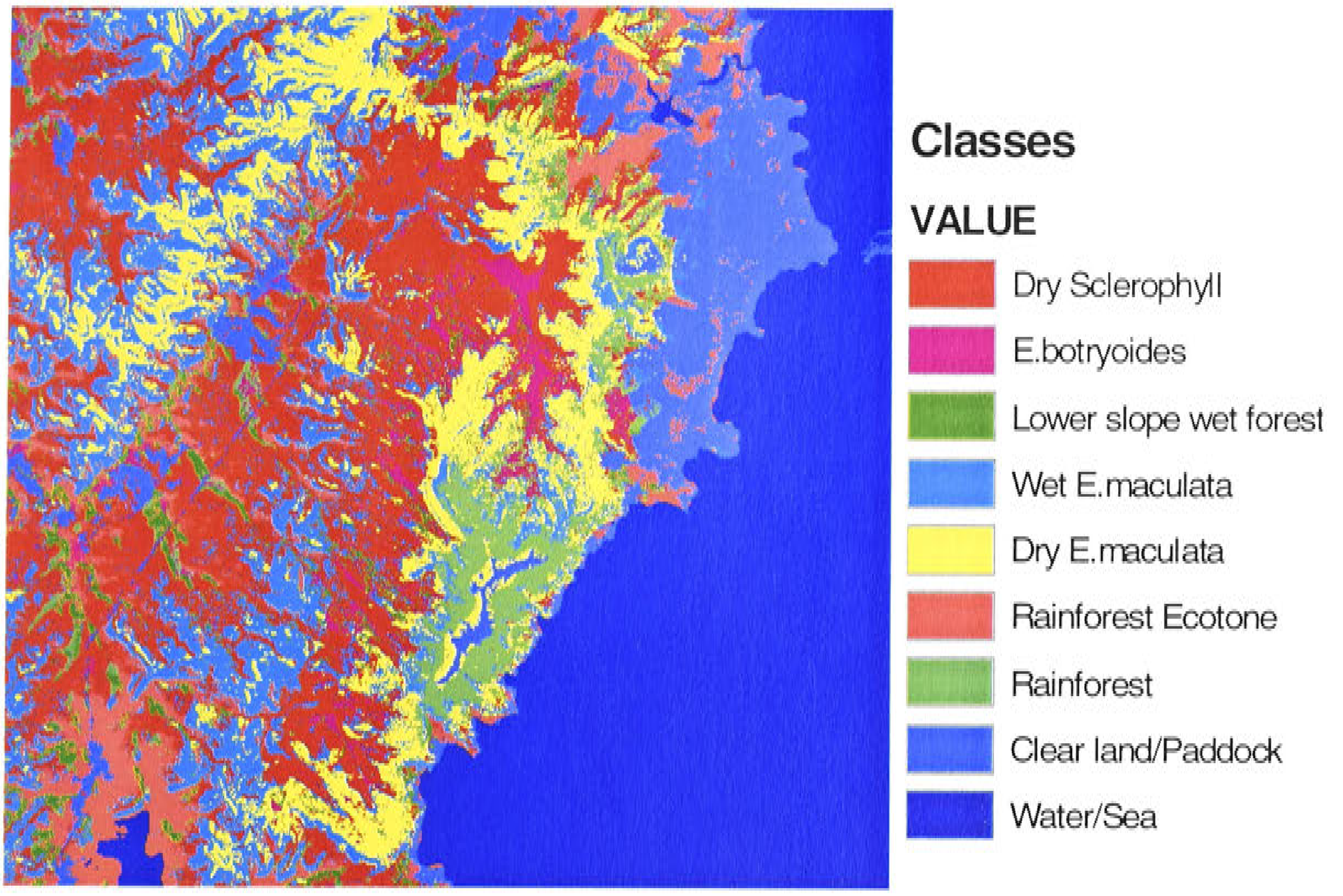


Plate 9 Classification map of maxclass

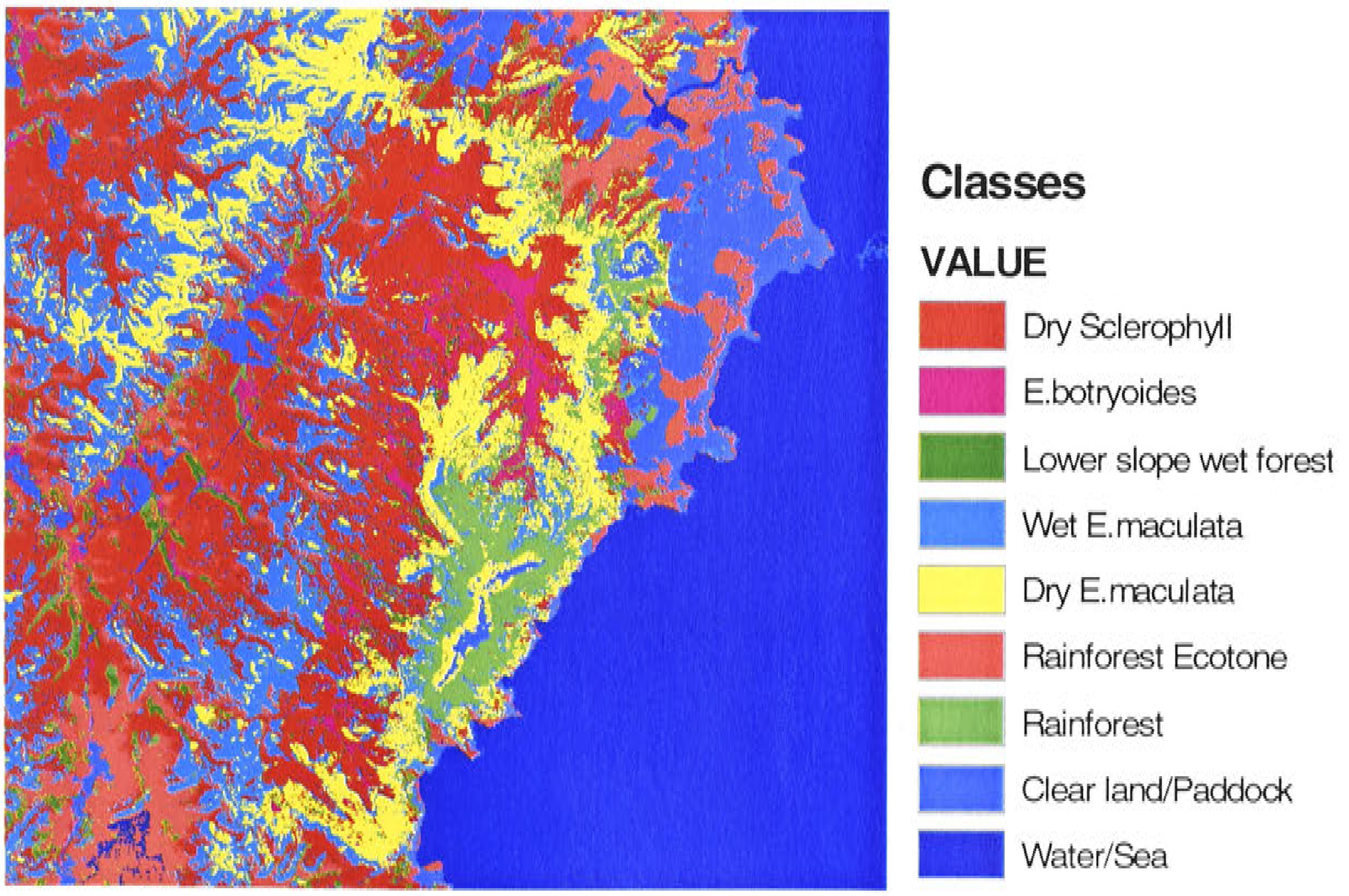


Plate 10 Classification map of ffamaxclass

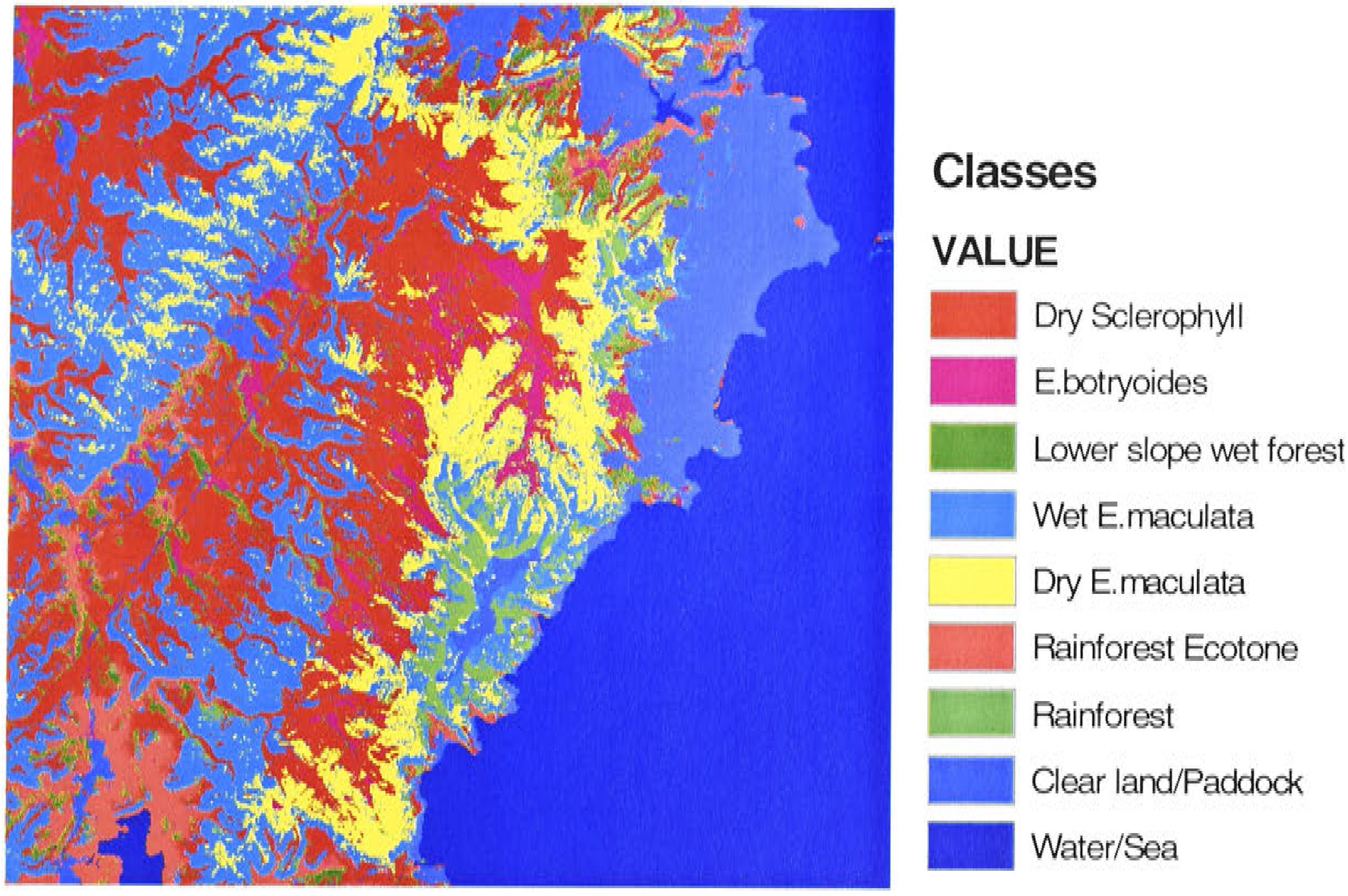


Plate 11 Classification map of fameanppclass

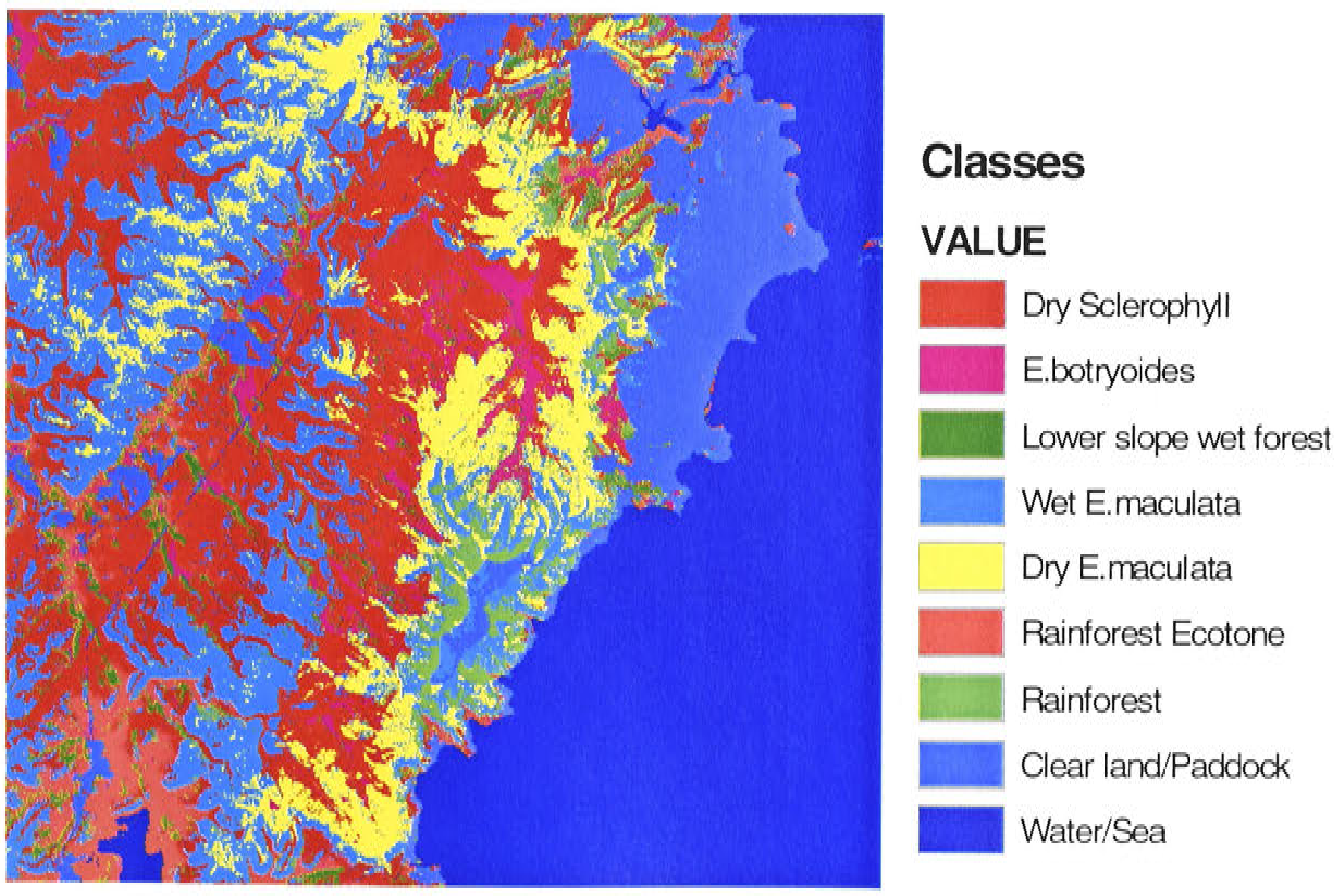


Plate 12 Classification map of fmeanclass

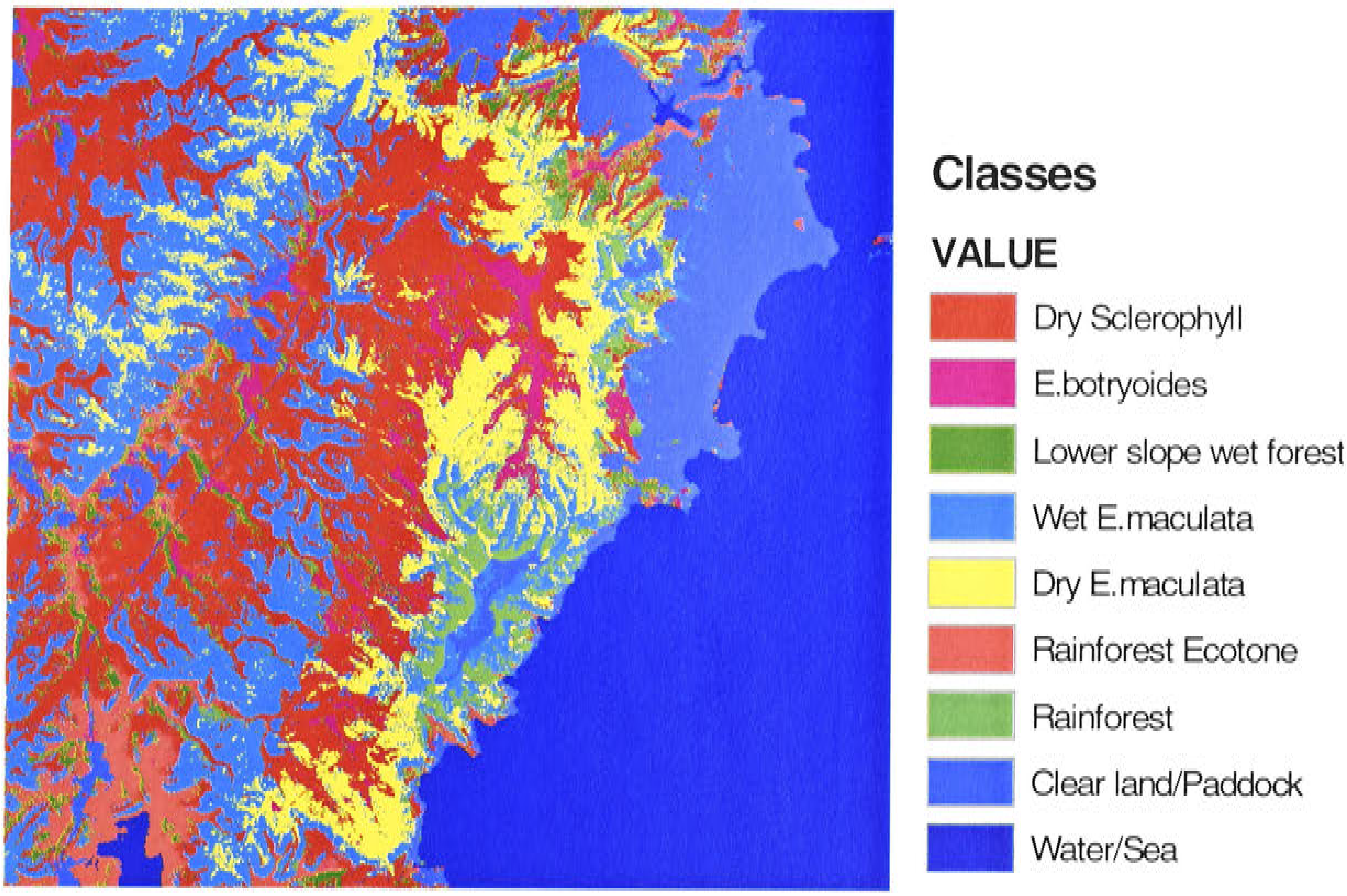


Plate 13 Classification map of fmeanppclass

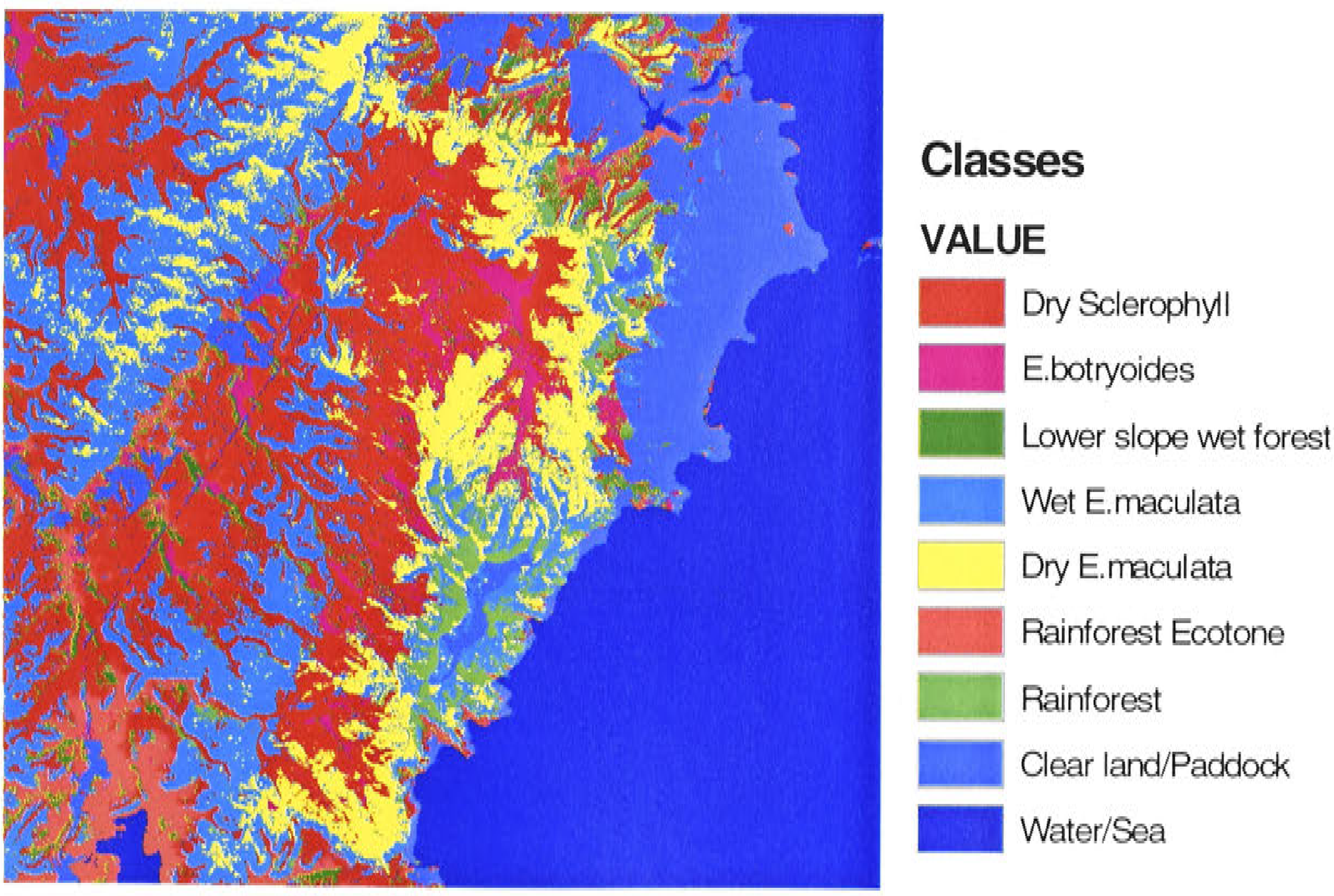


Plate 14 Classification map of fameanclass

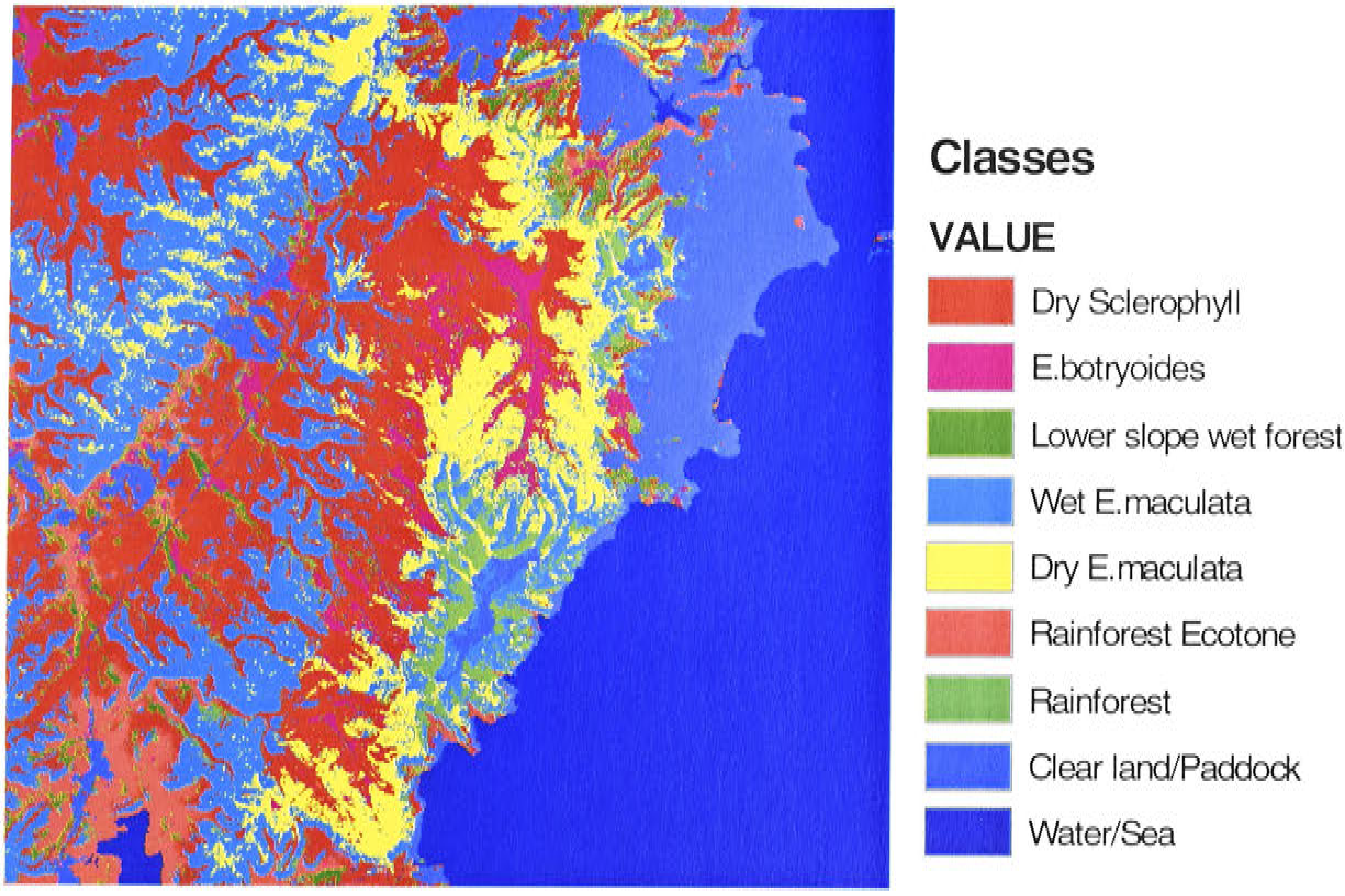


Plate 15 Classification map of fmedianclass

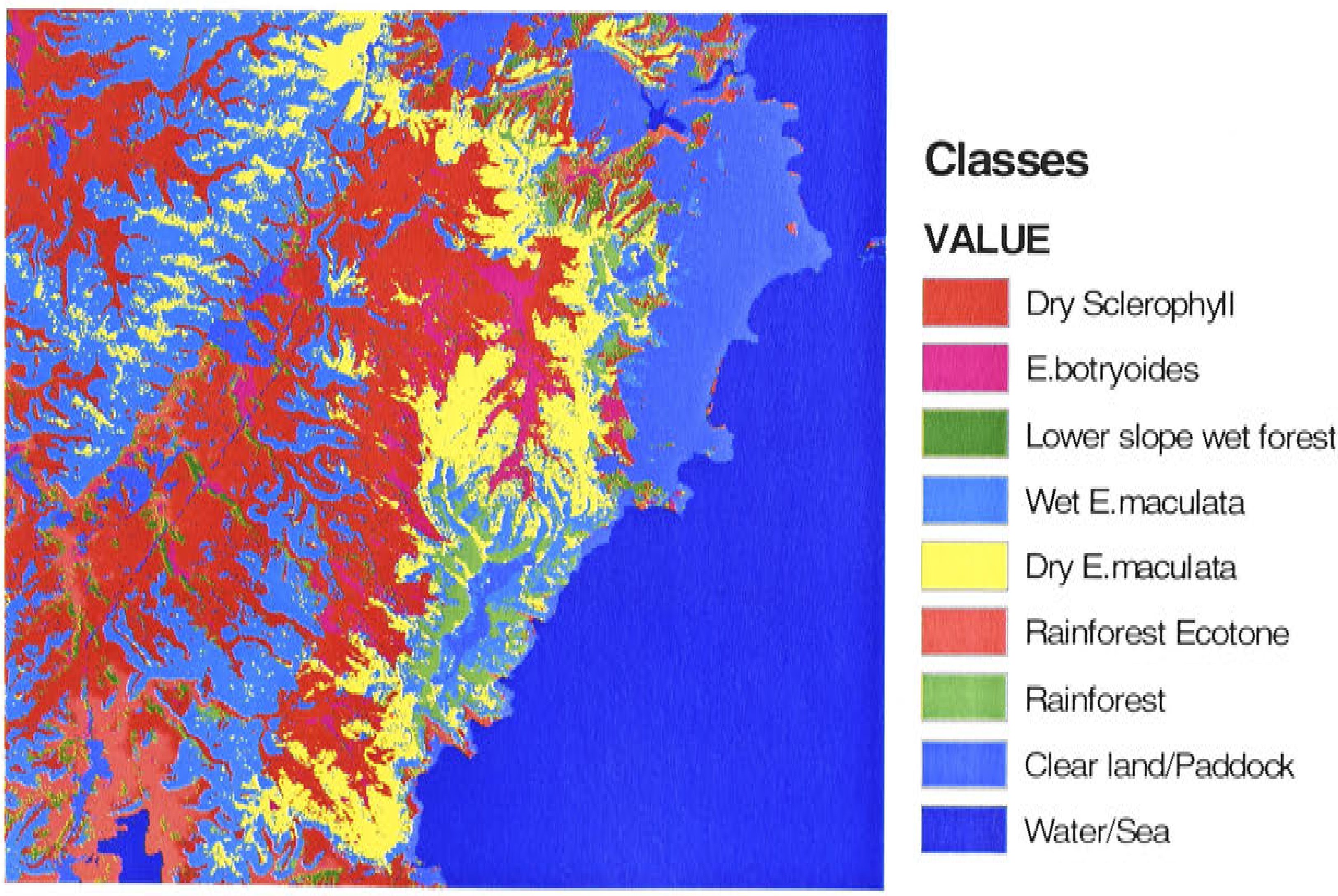


Plate 16 Classification map of fmedppclass

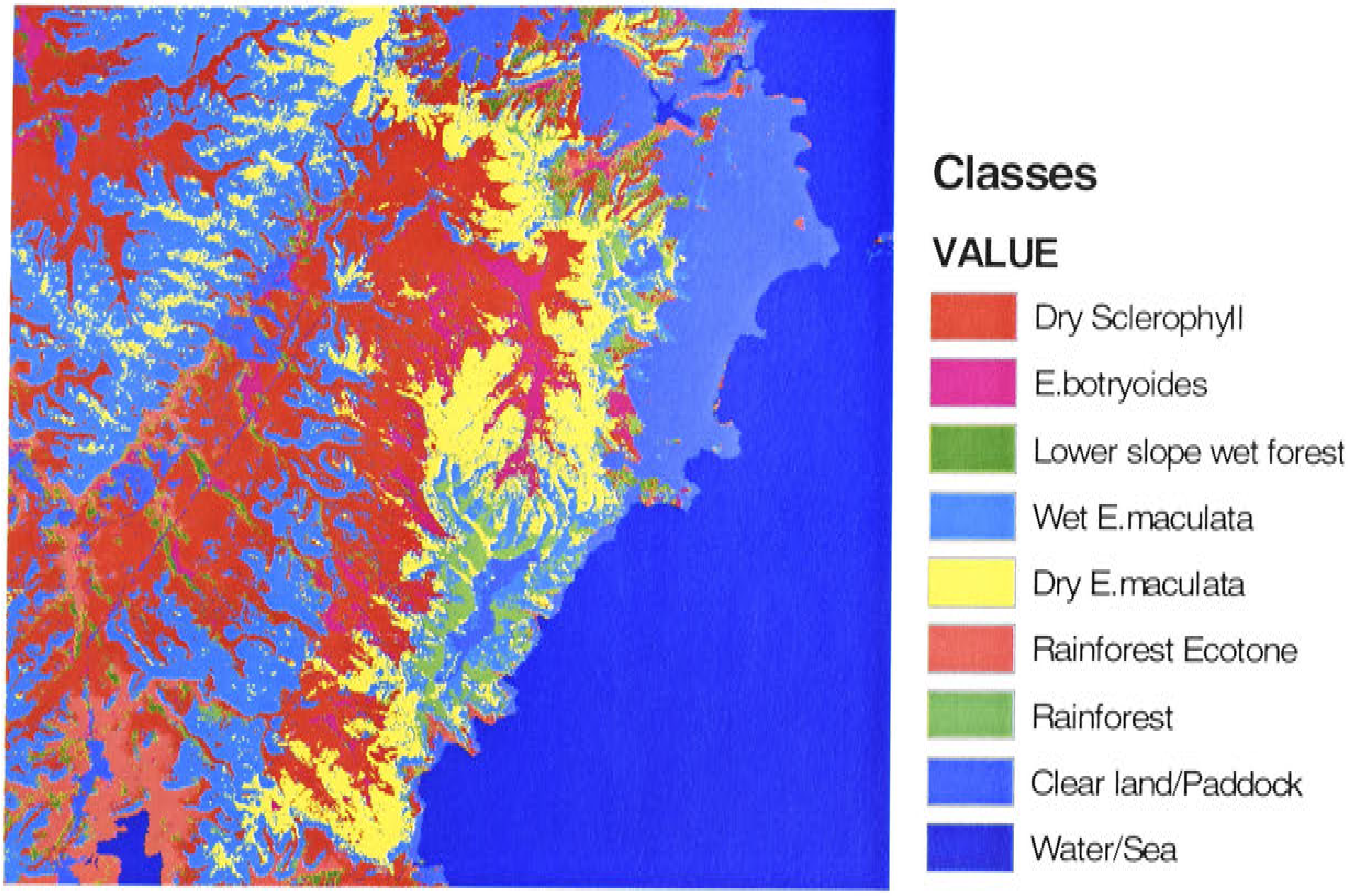


Plate 17 Classification map of famedianclass

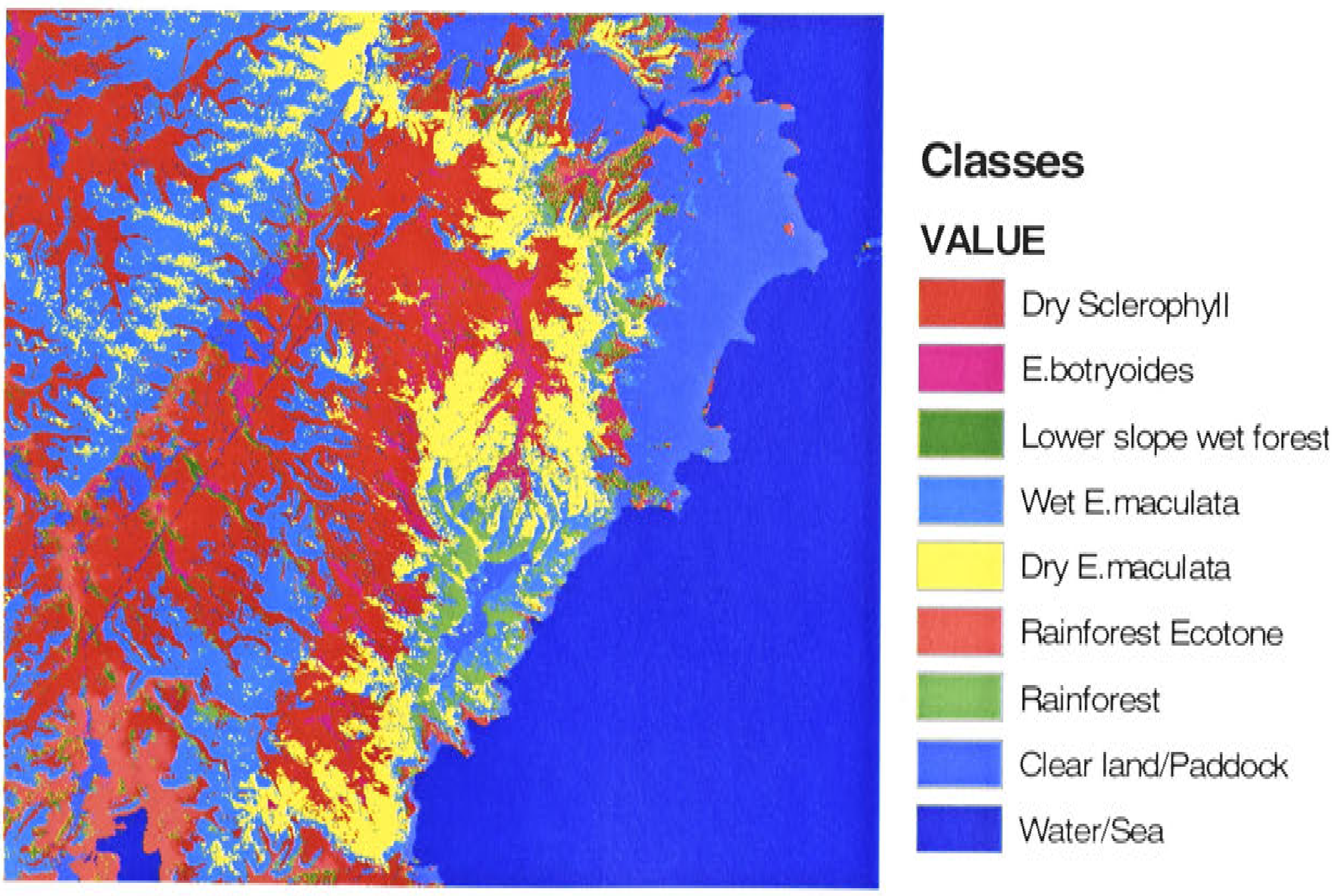


Plate 18 Classification map of famedppclass

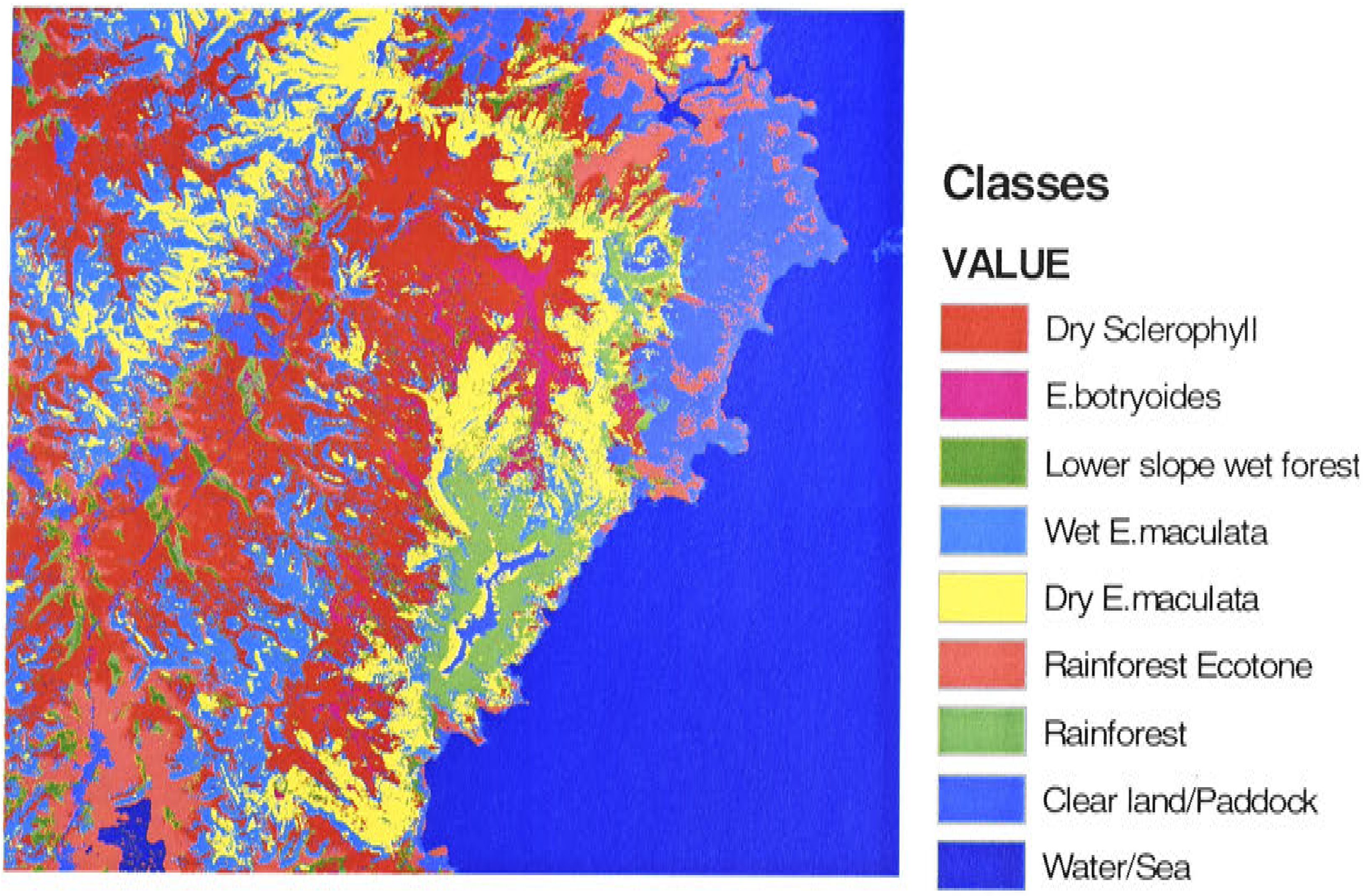


Plate 19 Classification map of fmaxclass

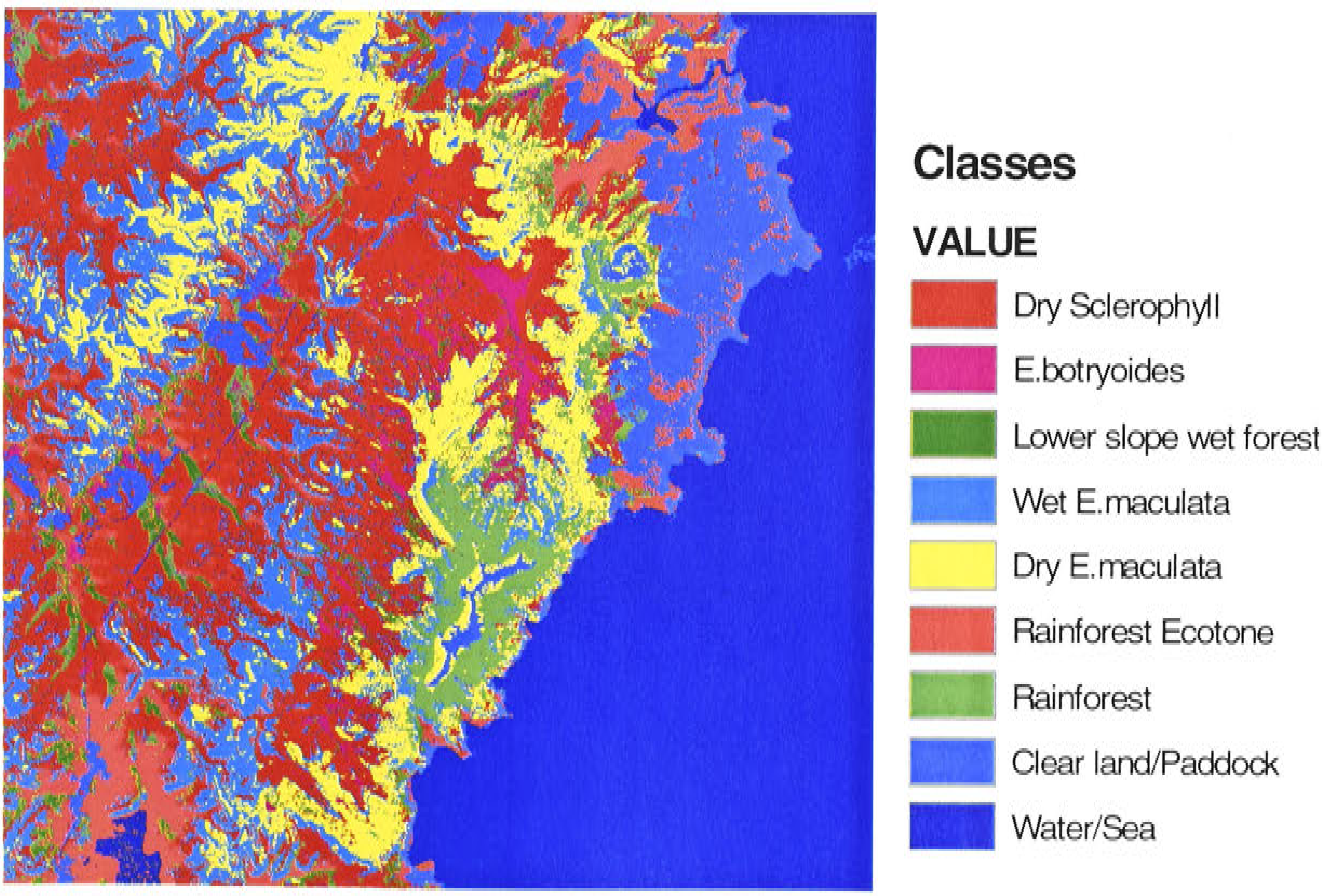


Plate 20 Classification map of fmaxppclass

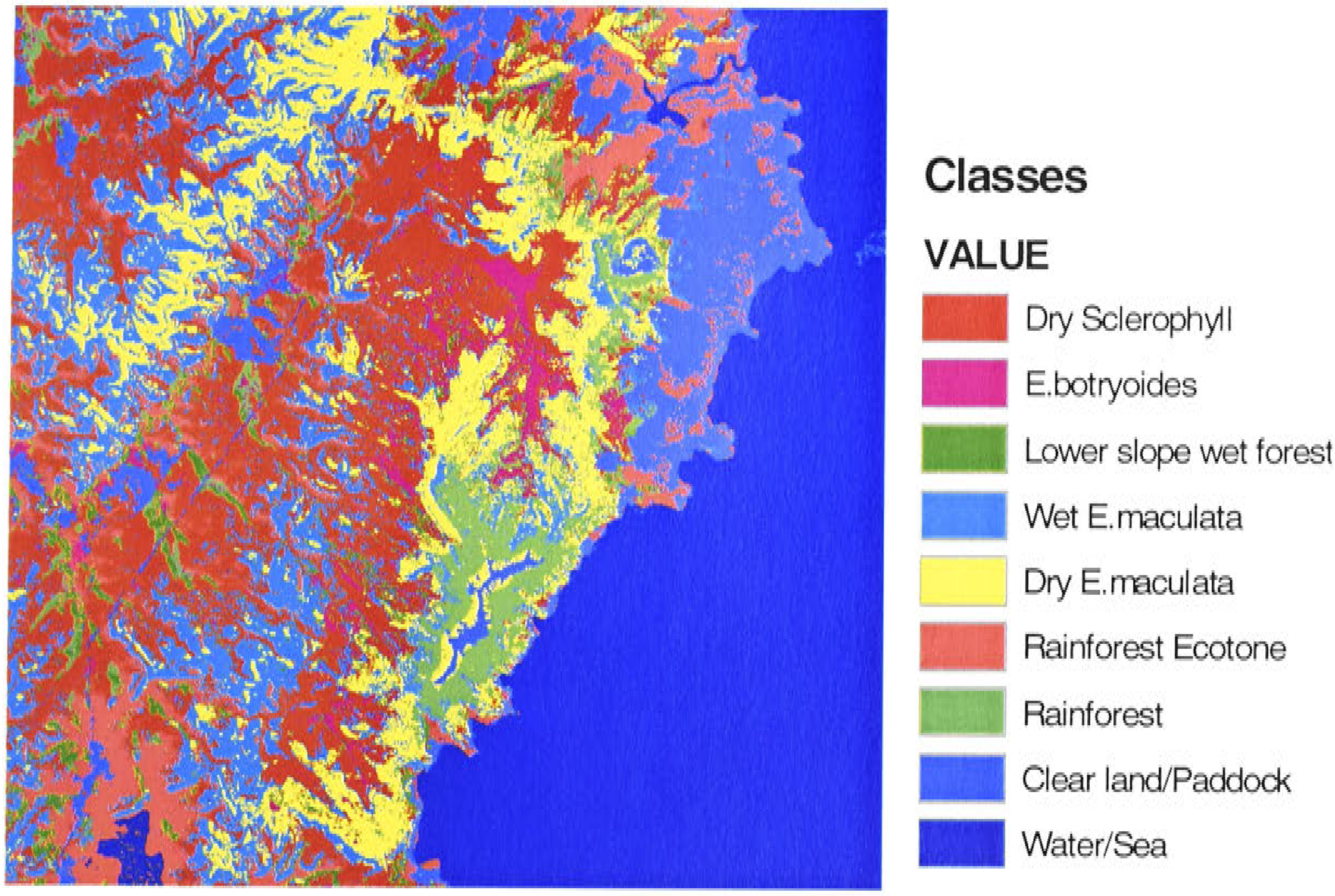


Plate 21 Classification map of famaxclass

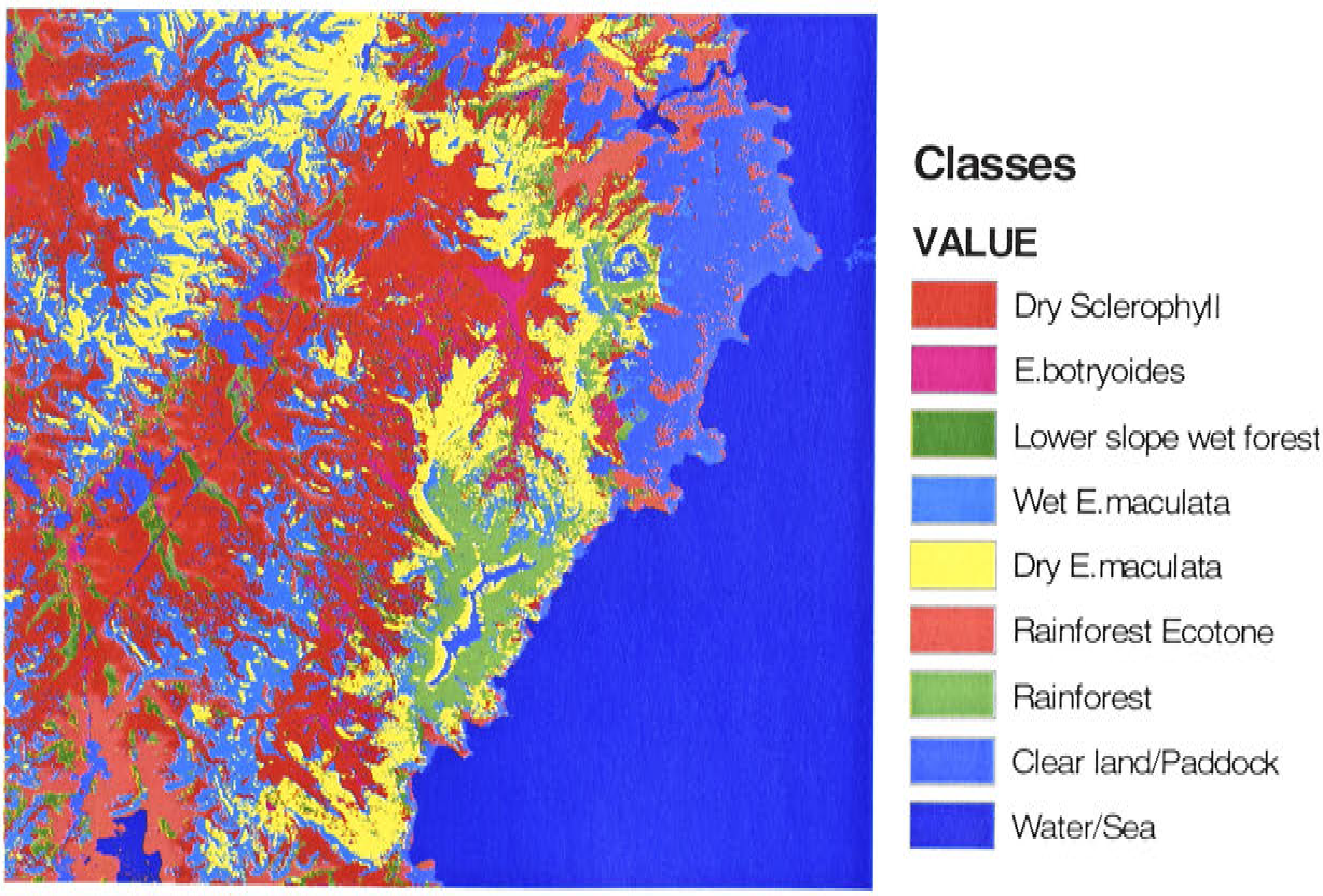


Plate 22 Classification map of famaxpclass

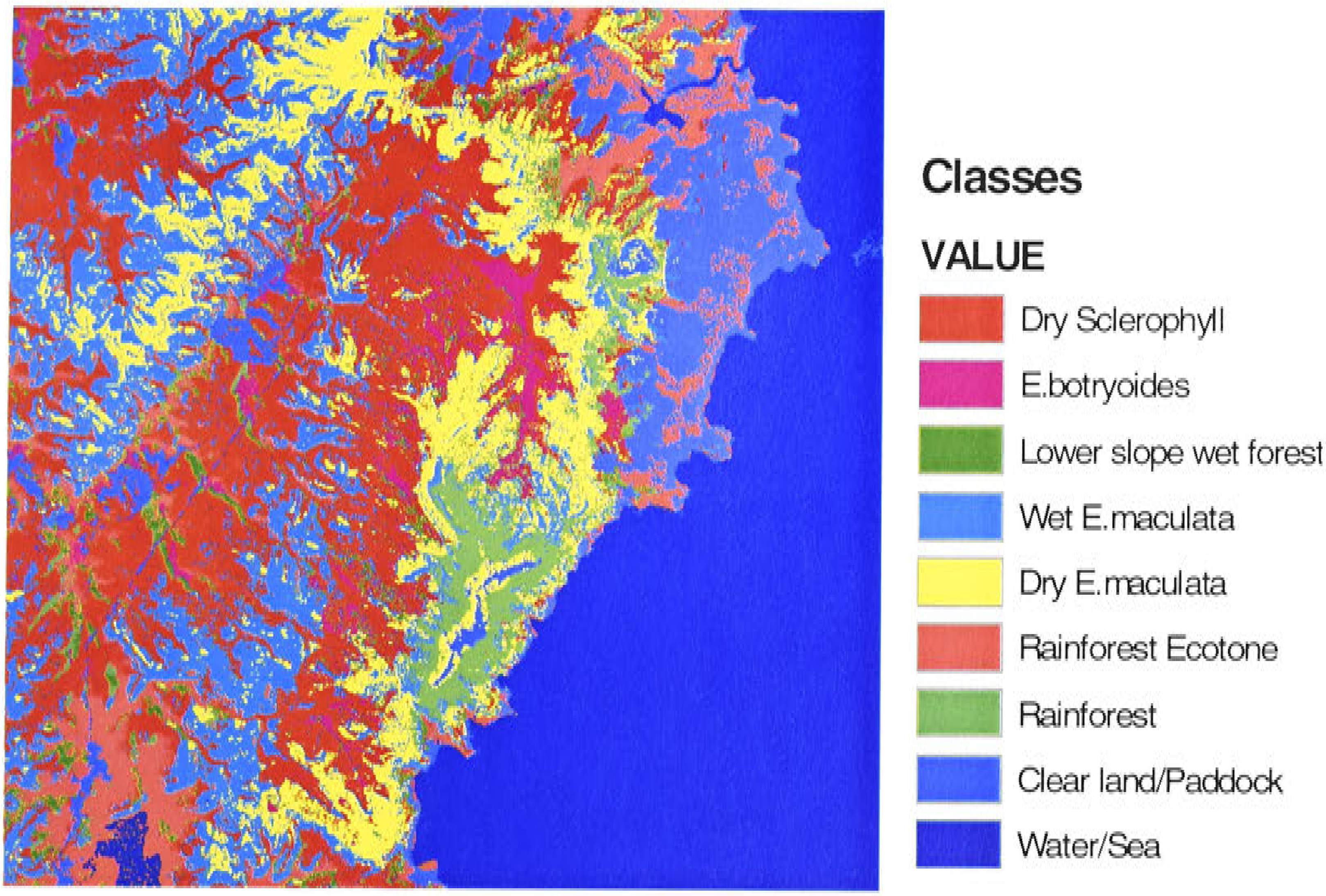


Plate 23 Classification map of fcmaxclass

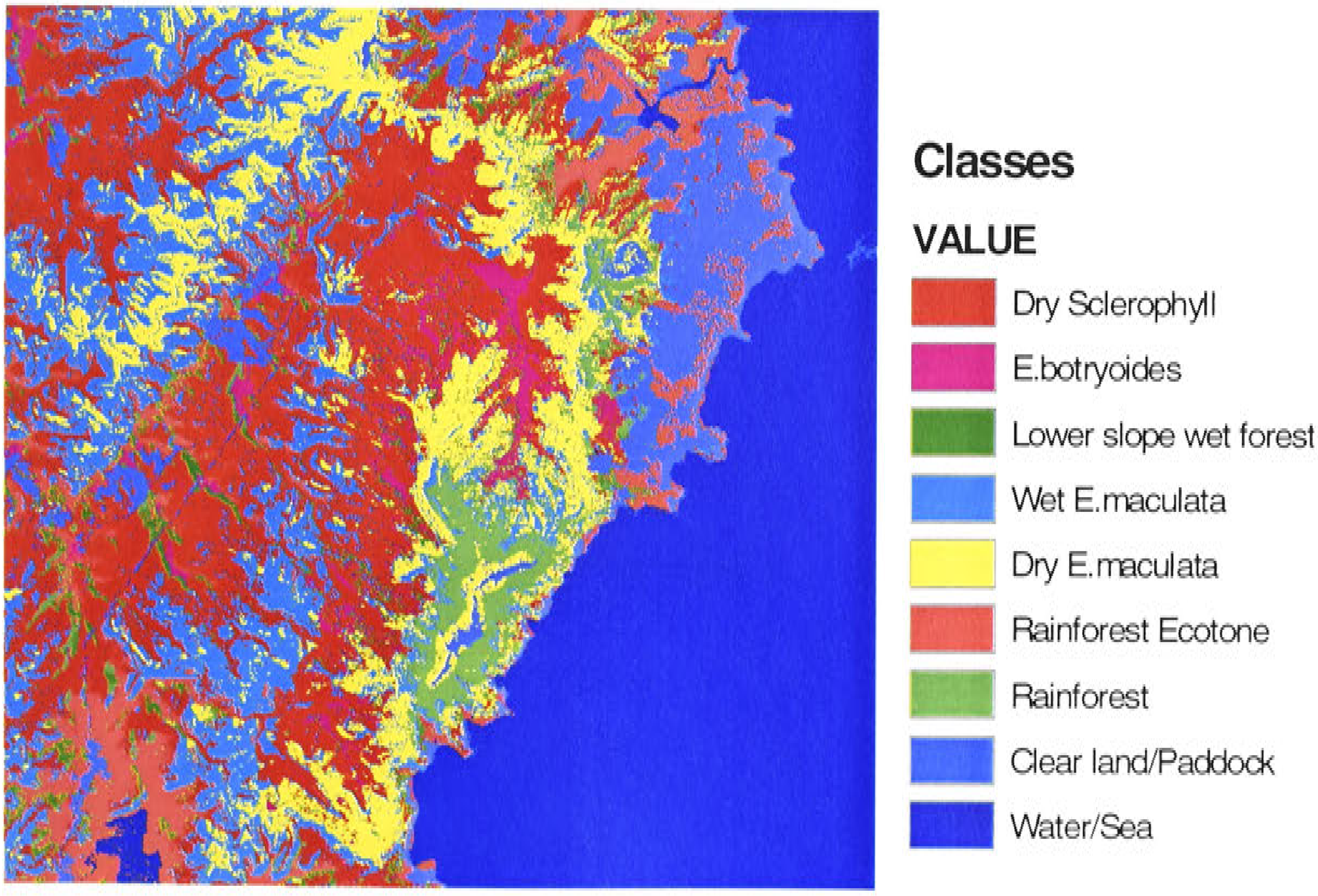


Plate 24 Classification map of fffmaxclass

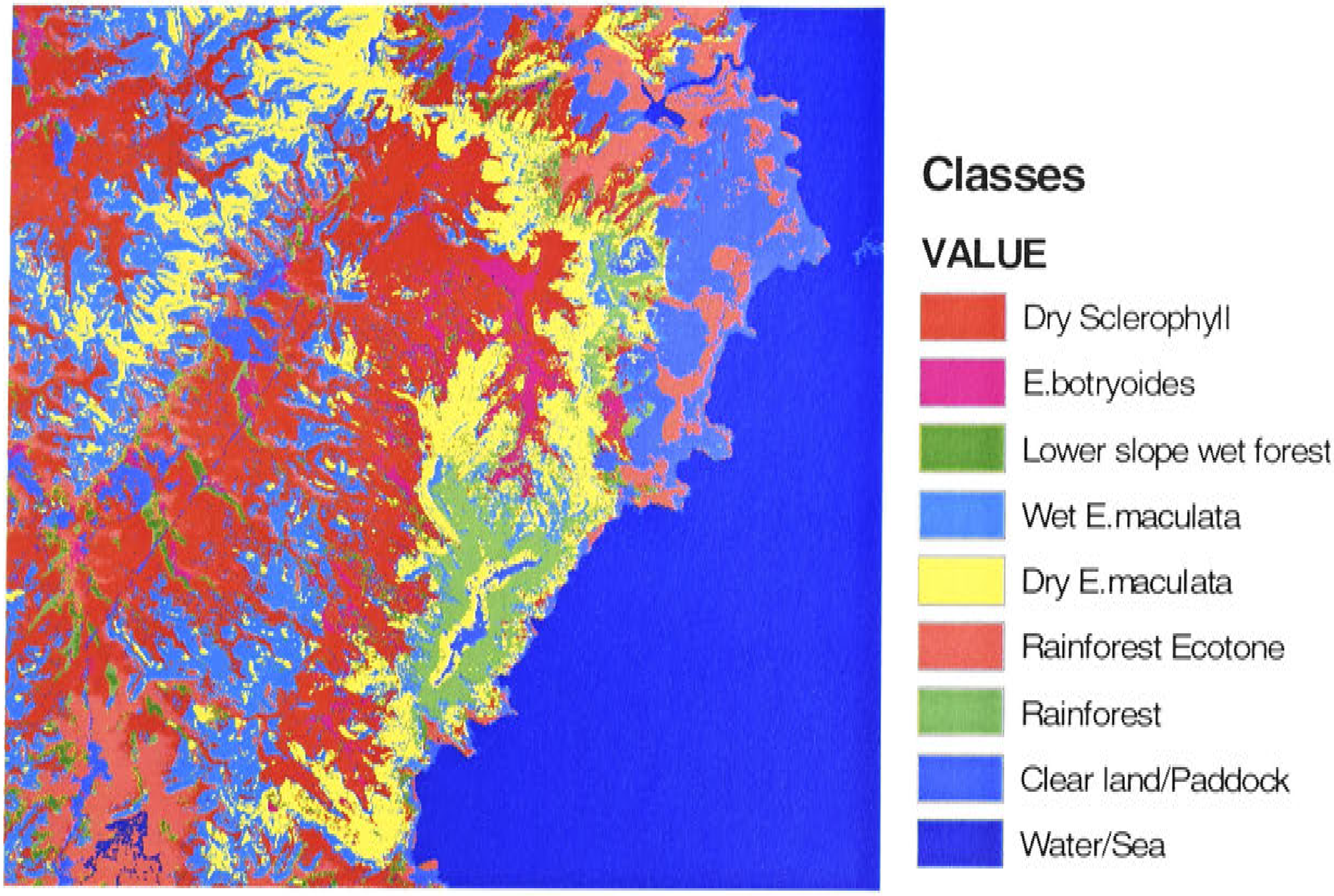


Plate 25 Classification map of fffamaxclass

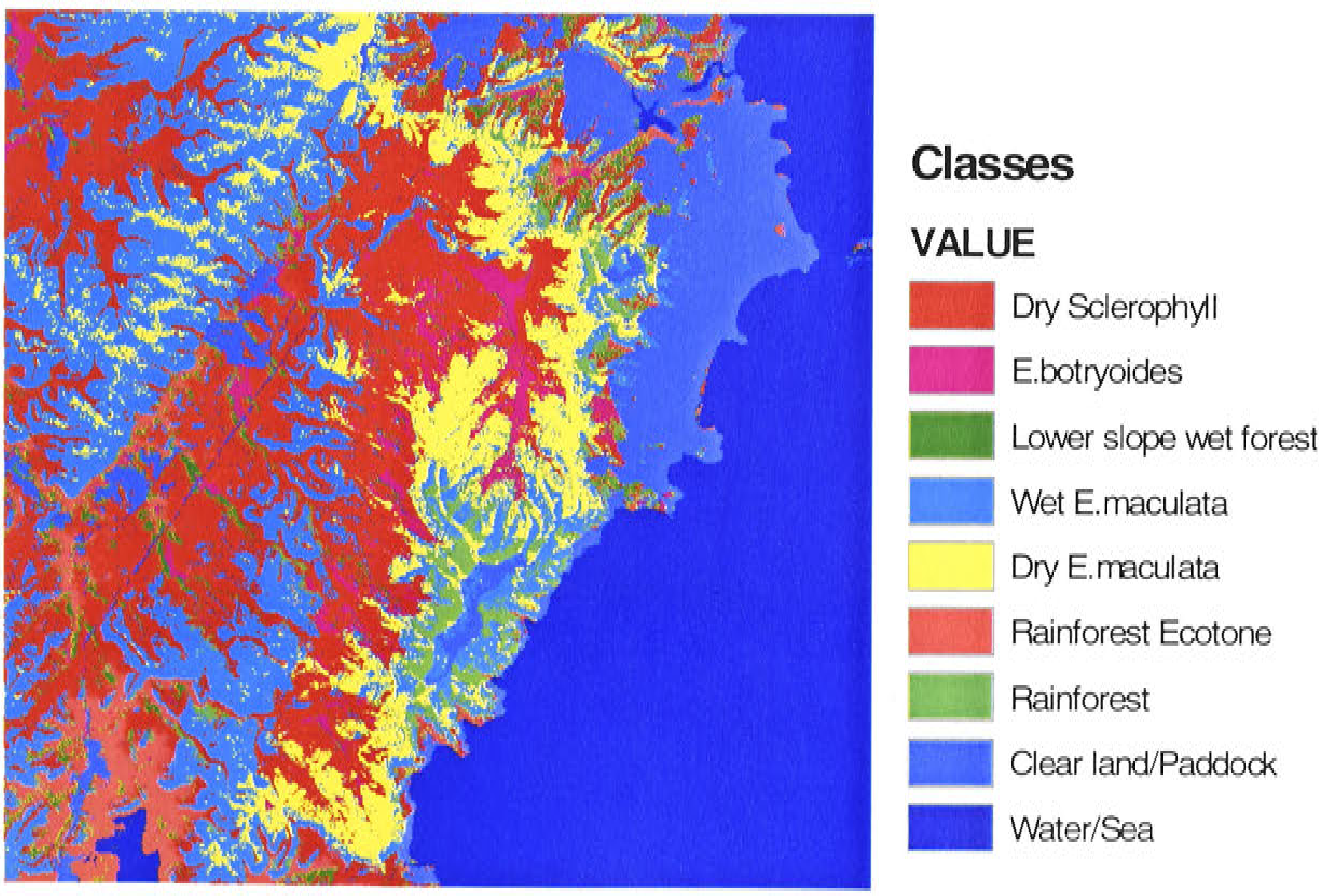


Plate 26 Classification map of ffagreeclass

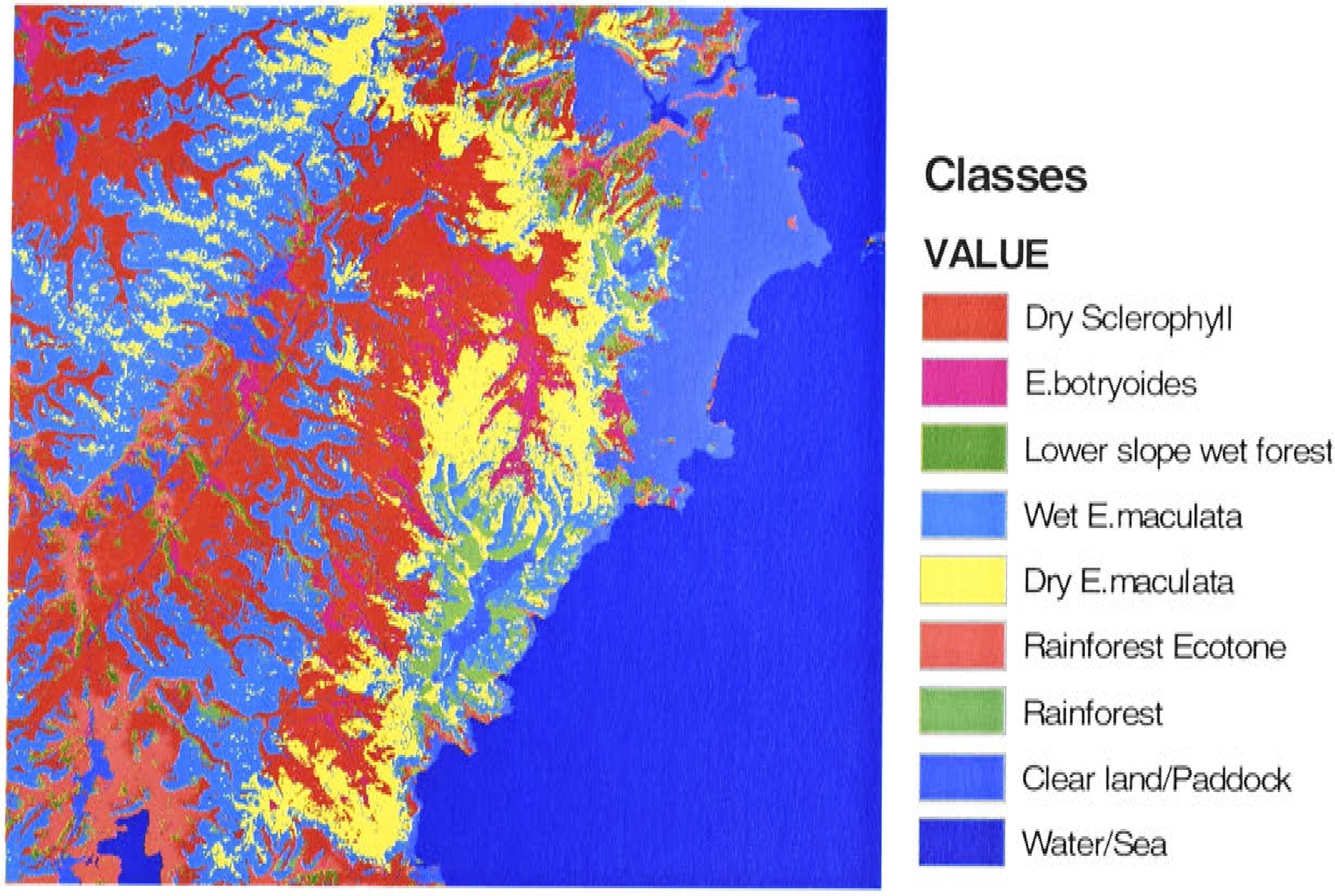


Plate 27 Classification map of ffappclass

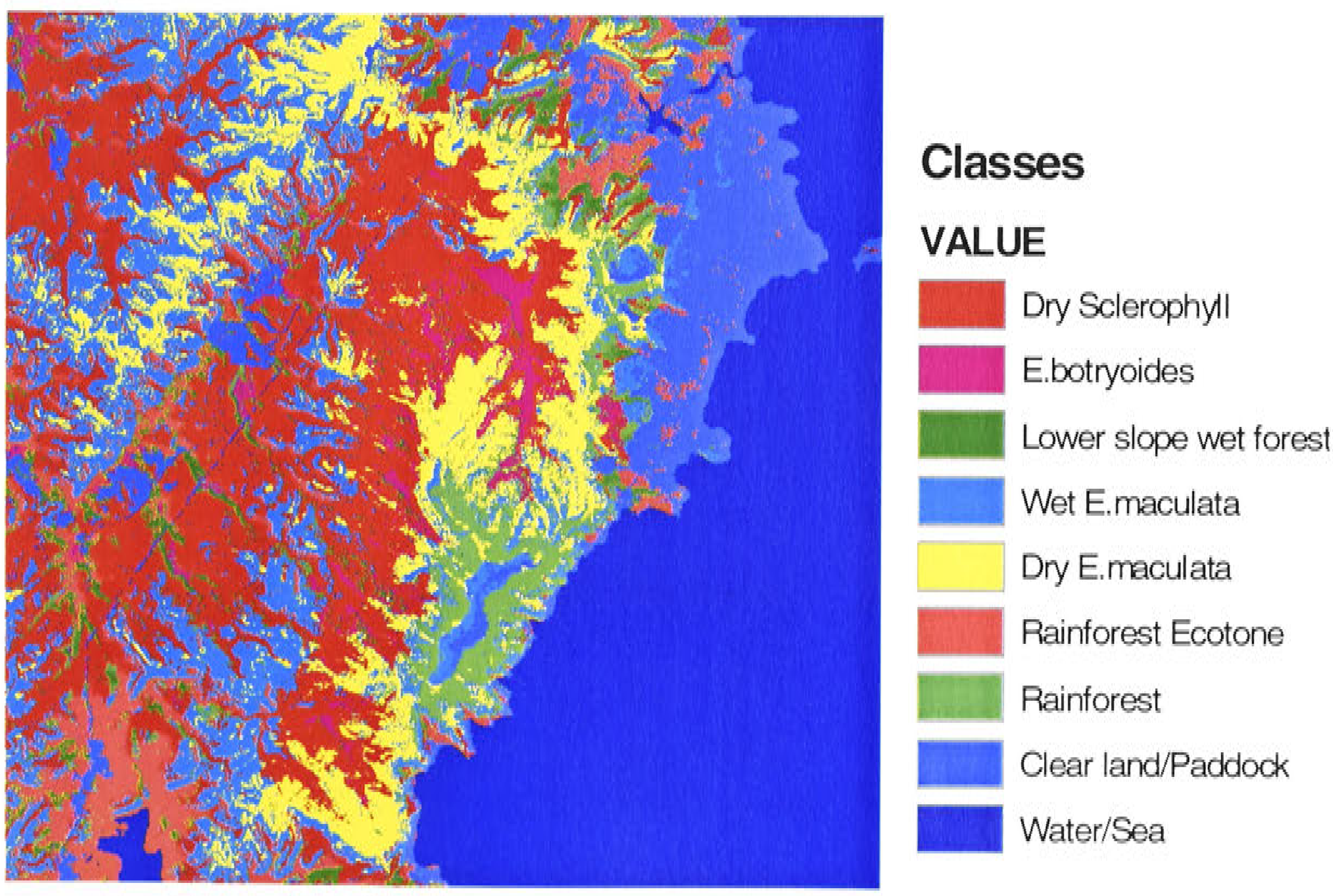


Plate 28 Classification map of vote1

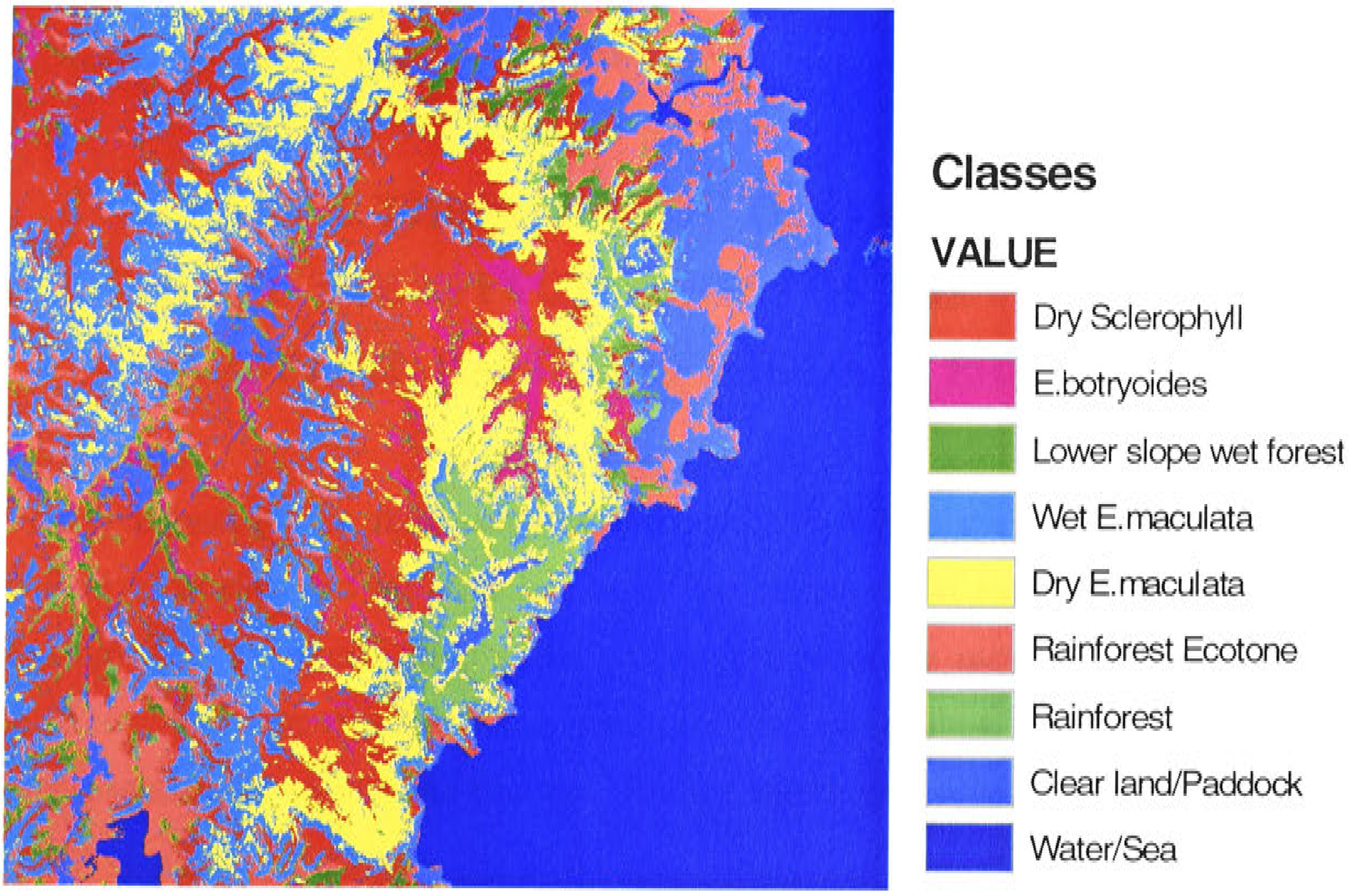


Plate 29 Classification map of vote6

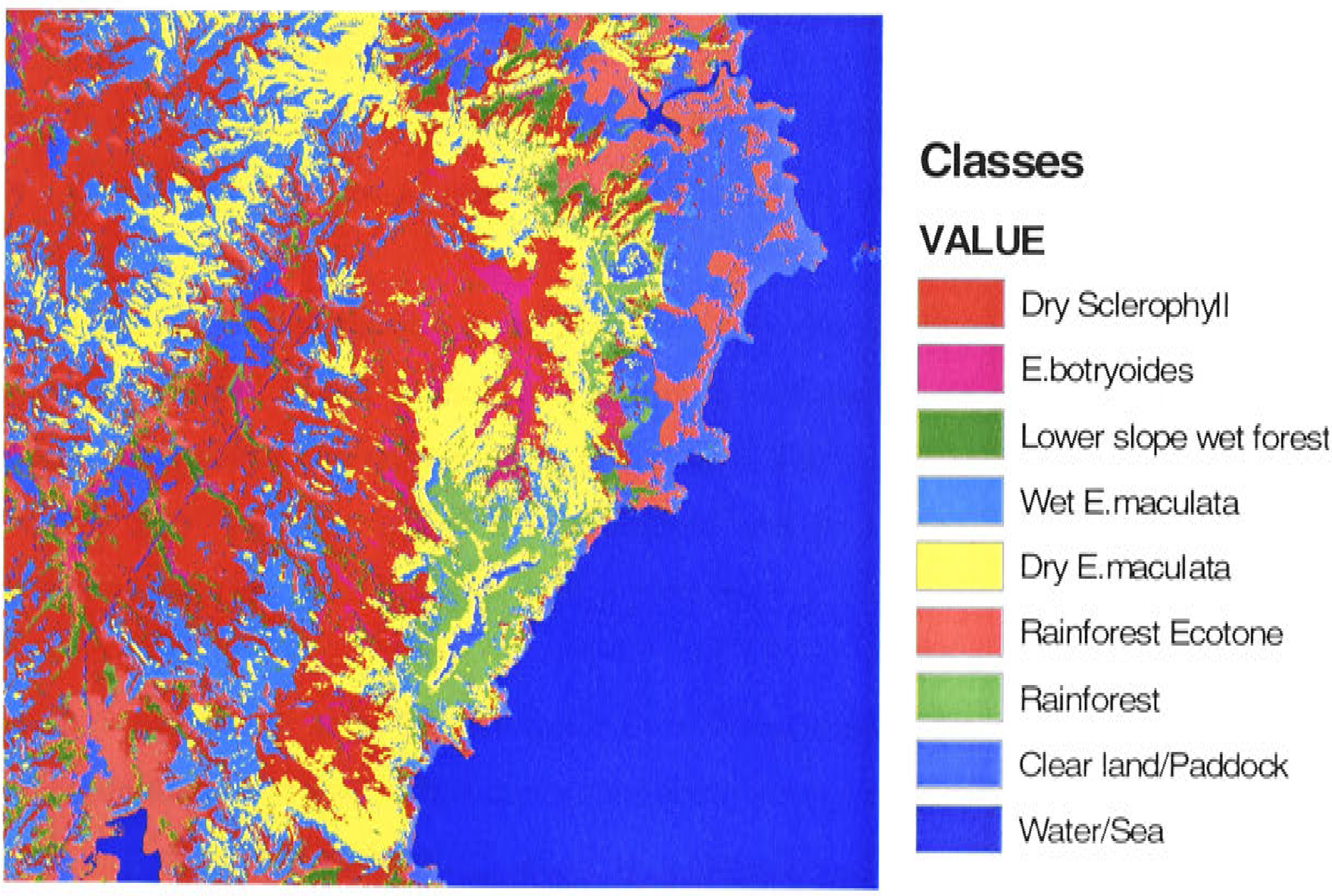


Plate 30 Classification map of vote7

Chapter 7

EVALUATING THE MODES OF THE INDIVIVIDUAL AI MODELS AND THE EFFECTIVENESS OF THE COMBINED AI MODELS IN HANDLING THE DATA ERRORS

The chapter presents the third stage of this study, in which the modes of the three individual AI models and the effectiveness of the combined AI models in handling the data errors were examined. Firstly, the chapter describes the methods applied in this stage. Then, the chapter reports the results of the evaluations. Following that, the chapter discusses the results and findings. Finally, the chapter gives a short summary of this stage of study.

7.1 METHODS

After carefully examining the data sources, two apparent data errors have been identified in this study (see section 3.2). Both data errors are located on Brush Island area (Figure 3.1). This study wants to answer the following questions:

- Did the individual AI models handle the data errors differently?
- If so, why did the individual AI models handle the data errors differently?
- Could the combined AI models handle the data errors satisfactorily?

To answer the first and the last questions, the study did not intend to use any kind of error index; instead visual assessments of the classification maps from these models were used. The focus was on the Brush Island area. The visual assessment was done by looking at how well the island was predicted and how many error samples were corrected by each model.

Because the available error models (both formal mathematical models and simulation models) are not applicable in this study (see section 1.4), monitoring error propagation through the modeling process in detail is impossible. Thus, to answer the second question, a method was developed to gain better insight into the modeling processes. The method is to add independent input variables into the modeling process one by one and evaluate the classification results. The visual assessment method described above was again used. At the same time, the correctly classified forest samples were counted in the process to see whether or not adding independent variables would increase the predictive accuracy of the individual models.

7.2 RESULTS

The following table (Table 7.1) gives the answers to the above first and last questions by examining the classification maps of the individual and combined AI models (Plates 1-30).

Table 7.1 Summary of how Brush Island was predicted and how many error samples out of four were corrected by the individual and combined AI models

Name of the model	How Brush Island was predicted?	How many samples were corrected?	Name of the model	How Brush Island was predicted?	How many samples are corrected?
Decision Tree	No	0	Artificial Neural Network	Part	1
D-S	Perfect	4	Combine1	Small part	0
Combine2	Small part	0	D-S1	Very good	3
meanclass	Part	0	medianclass	Part	0
maxclass	Large part	2	ffamaxclass	Large part	3
fmeanppclass	Part	1	fameanclass	Part	0
fameanppclass	Part	0	fmeanclass	Part	1
fmedianclass	Part	0	fmedppclass	Part	0
famedianclass	Part	0	famedppclass	Part	1
fmaxclass	Large part	2	fmaxppclass	Large part	2
famaxclass	Large part	2	famaxpclass	Large part	2
fcmaxclass	Large part	3	fffmaxclass	Large part	3
fffamaxclass	Large part	3	ffagreeclass	Part	1
ffappclass	Part	1	Vote1	Very good	3
Vote6	Large part	3	Vote7	Large part	3

Visually comparing the three initial classifications indicates that the Decision Tree did not predict Brush Island. The Artificial Neural Network delineates part of the island boundary and could correct only 1 out of 4 error samples. However, the model based on Dempster-Shafer's theory did perfectly distinguish the island, and all of the four error samples were successfully corrected. At the first stage combination, the two combined AI models based on the majority voting system (combine1 and combine2) have only predicted a small part of Brush Island and could not correct a single sample from error. The combined AI model of D-S1 based on Dempster's rule of combination, on the other hand, has nicely predicted the island, and 3 out of 4 error samples have been corrected. Among the three combined AI models based on simple statistical functions, meanclass and medianclass could only partly predict the island, and none of the error samples has been corrected. Maxclass, however, has predicted the island in a large part and corrected 2 out of 4 error samples. The results of the eighteen combined AI models based on fuzzy set theory are variable. Those models based on the measurements of difference of bmeanMs, bmedianMs, and bMs have predicted only part of the island, and they could correct at most 1 out of 4 error samples. Meanwhile, the four combined AI models based on bmaxMs have consistently predicted the island in a large part and corrected 2 out of 4 error samples. The four combined AI models based on bmaxCs have also predicted the island in a large part, and they have successfully corrected 3 out of 4 error samples. At the second stage combination, all of the three combined AI models based on the majority voting system have corrected 3 out of 4 error samples. While vote6 and vote7 have predicted the island in a large part, vote1 has predicted the island better.

Figures 7.1-3 give the answers to the second question. Figure 7.1 shows how the known data errors manifested through the Decision Tree model by adding independent variables into the model step by step. When only the three Landsat TM bands were used as input variables to predict the forest type classification, the Decision Tree correctly classified 72 out of 228 forest type samples, a small part of the island was predicted but none of the error samples was corrected. When the aspect variable was added to the Decision Tree model, 68 forest samples were correctly predicted, the island shape improved, and 1 of 4 error samples was corrected. When the DTM variable was further added, 85 forest samples were correctly predicted and the island was discriminated very well. As well, 3 of 4 error samples were corrected. When the slope variable was added, 99 forest samples were correctly predicted, the island shape remained good, and 3 of 4 error samples were corrected. However, when the geology variable was finally added, though the correctly predicted forest samples increased to 106, the island completely disappeared, and none of the 4 error samples were corrected.

Figure 7.2 shows how the known data errors manifested through the Artificial Neural Network model by adding independent variables into the model step by step. For the Artificial Neural Network, when only the three Landsat TM bands were used for the prediction of forest types, 66 forest samples were correctly predicted, the island shape was good, and 2 of 4 error samples were corrected. When the aspect variable was added, 75 forest samples were correctly predicted, the island was well discriminated, and 2 of 4 error samples were corrected. When the DTM variable was added, 76 forest samples were correctly predicted, the island was well described, and 2 of 4 error samples were corrected. When the slope variable was added, the correctly classified forest samples increased to 93, the island northern boundary was wrongly extended, and 2 of 4 error samples were corrected. When the geology variable was added, 102 forest samples were correctly classified, the island shape was worse, and only 1 of 4 error samples was corrected. This result for the Artificial Neural Network is slightly different from that of the section 5.1 because randomising the initial weights for the Artificial Neural Network caused different results for this subsequent run.

Figure 7.3 shows how the known data errors manifested through the model based on Dempster-Shafer's theory by adding independent variables into the model step by step. For example, when only the three Landsat TM bands were used, only 35 forest samples were correctly classified, but the island boundary was well described, and 3 of 4 error samples were corrected. When the DTM variable was added, the correctly predicted forest samples increased to 79. Not much change to the island boundary was found, and 3 of 4 error samples were corrected. When the slope variable was added, 88 forest samples were correctly classified, the island boundary looks better, and all of the 4 error samples were corrected. But when the geology variable was added, the correctly classified forest samples decreased to 85, the predicted island shape was worse, and the correction of error samples decreased to 3 out of 4. When the last input variable, aspect, was added the correctly classified forest samples increased to 89, the island boundary returned to its previous state, and all 4 error samples were successfully corrected.

7.3 DISCUSSION

This stage of study has helped us to gain some insight into the modes of the three AI models in handling the known data errors, even though it did not apply a formal error model. The combined effect of the data errors associated with the geology variable and the sampling error in the Brush Island area was manifest differently through the modelling processes of the three models. The Decision Tree was the most affected with the island being lost and the error samples not being corrected. The Artificial Neural Network was partially affected with the island only being partly shown and one error sample being corrected. The model based on Dempster-Shafer's theory was not affected by the known error, for the island was perfectly predicted and all error samples were successfully corrected.

Adding input variables one by one into the three individual models did help to understand how the known input errors have been manifested through the three models. It has clearly shown that the geology variable could increase the predictive accuracy (e.g., for the Decision Tree and the Artificial Neural Network), but definitely had a negative effect in predicting the island and in correcting the sampling error. However, the magnitude of this effect is variable, For the Decision Tree, the effect was critical, as after adding the geology variable the island completely disappeared. For the model based on Dempster-Shafer's theory, the effect is less critical as it slightly reduced the predictive accuracy, slightly worsened the shape of the island boundary, and decreased the correction of the error samples from 4 to 3. For the Artificial Neural Network, the effect was slightly greater than the model based on Dempster-Shafer's theory, as the island shape was worse than that of the model based on Dempster-Shafer's theory, and the correction of the error samples was decreased from 2 to 1.

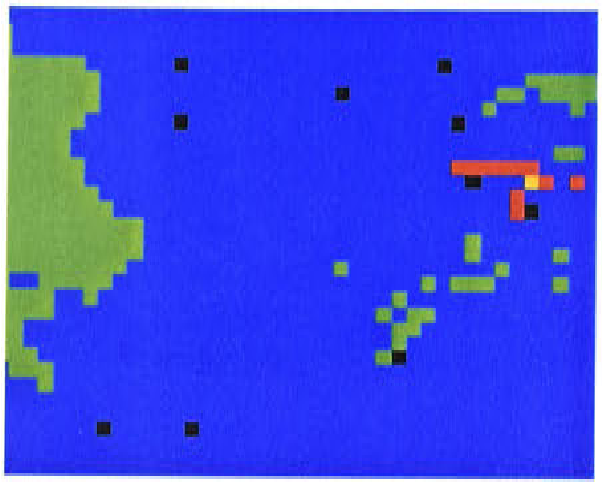
The reasons for these differences lie in the different principles the three AI models are based on. Checking the resultant decision tree structure (Figure 5.1) it is found that the geology variable appears at the top of the tree (i.e., root), and the Water/Sea class was determined only by the geology variable. This obviously has a dominant influence on the classification result, which has misclassified Brush Island into the Water/Sea class, because the geology variable does not distinguish Brush Island from the surrounding sea. Therefore, it is understandable that the error associated with the geology variable would have had a large negative effect on the final product of the Decision Tree. On the other hand, Dempster-Shafer's theory assumes the input variables are independent and no one variable is able to dominate the others. So the negative effect of the geology variable was minor, and it has been compensated for by other input variables in the island area. For the Artificial Neural Network, it is more complicated. The connection weights are too many to be completely evaluated. Even the environmental variables have played a more important role in the classification than the remotely sensed data. It definitely was not dominated by any single input variable such as geology. At the same time, it is known that between the Decision Tree and the Artificial Neural Network, the Artificial Neural Network is the more effective user of samples, and is less sensitive to sampling errors than the Decision Tree (Lees, 1996c). Furthermore, it was also demonstrated that the model based on Dempster-Shafer's theory is the least sensitive model to sampling problems when compared with the Artificial Neural Network and the Decision Tree (see section 5.3). The above two points may well explain why the three models handled the sampling error differently.

One recommendation obtained from above discussion is that we should not take for granted the assertion that any AI model is more appropriate for handling data error. We need to spend more time in identifying error sources, understanding the principles of AI models, and analysing how error could be propagated through these models.

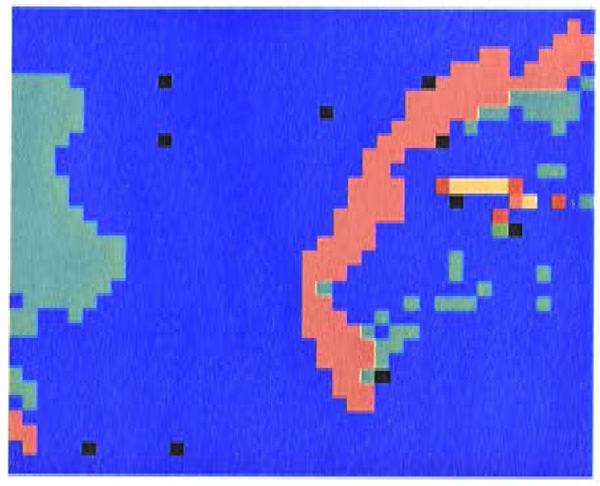
This study also strongly recommends using the combined AI models instead of using any individual AI models under uncertain application environments. While the combined AI models can improve classification performance, several of them such as D-S1, vote1, ffamaxclass, fcmaxclass, fffmaxclass, fffamaxclass, vote6, and vote7 have also done well in handling the data errors. It has been demonstrated that some combination approaches could inherit the good characteristics of error propagation mode from one model, but tend to suppress others that can accumulate the input error. For example, D-S1 largely inherited the advantage of the model based on Dempster-Shafer's theory in handling the known errors, while suppressing the bad results of the Decision Tree and the Artificial Neural Network.

7.4 SUMMARY

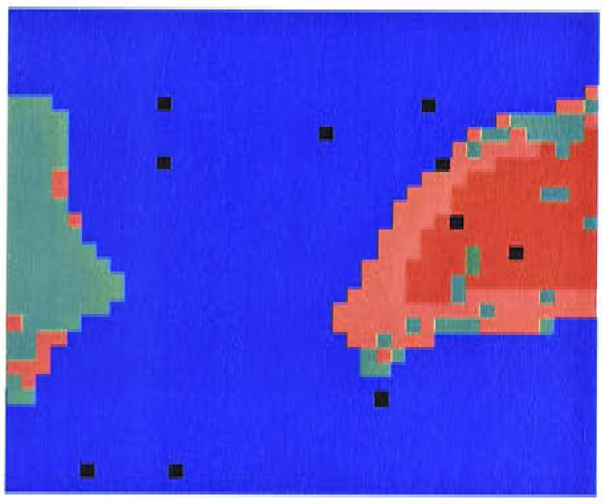
In summary, the author believes that different performances of the Decision Tree, the Artificial Neural Network and the model based on Dempster-Shafer's theory in handling the known data errors were due to their unique principles. It has been demonstrated that several of the combined AI models have effectively handled the known data errors. This further enhances the attractive features of the combination strategy. Therefore, if possible the combination strategy should be considered in the applications that have potential data errors and uncertainty.



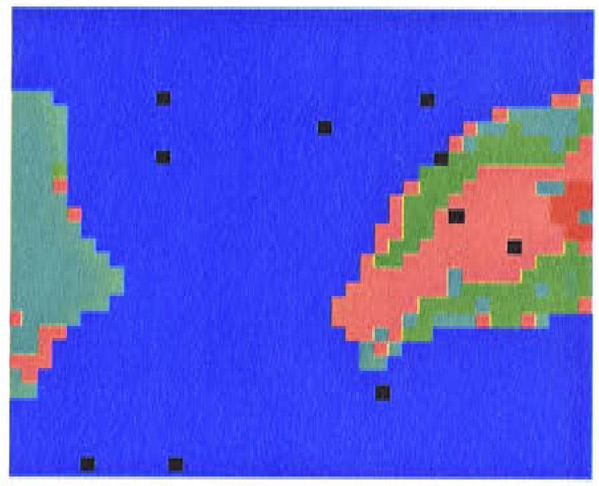
Landsat bands



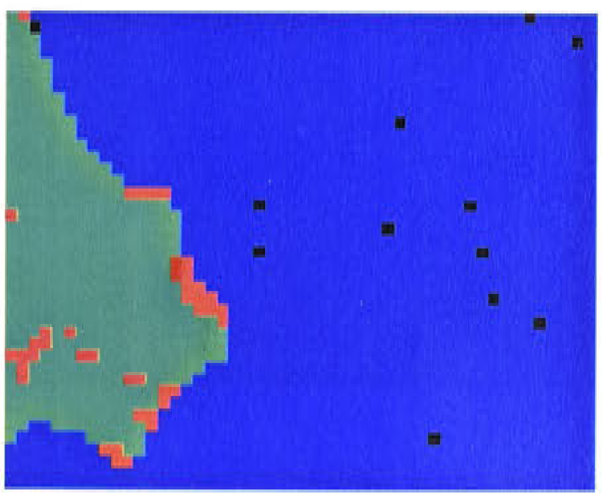
+ Aspect



+ DTM



+ Slope



+ Geology

Figure 7.1 Error manifestations through the Decision Tree by adding variables: from left to right, from top to bottom, Landsat bands, + aspect, + DTM, + slope, + geology

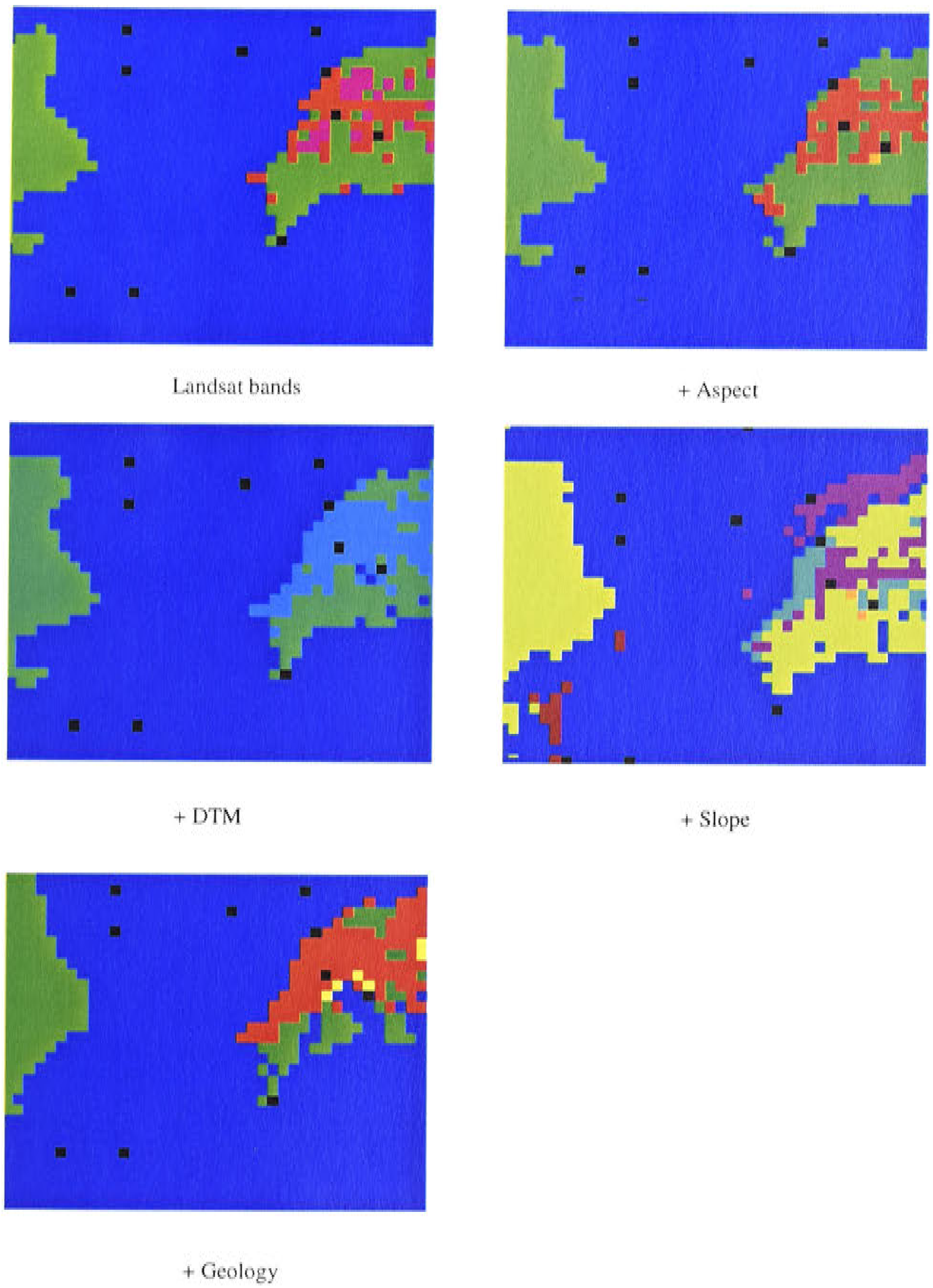


Figure 7.2 Error manifestations through the Artificial Neural Network by adding variables: from left to right, from top to bottom, Landsat bands, + aspect, + DTM, + slope, + geology

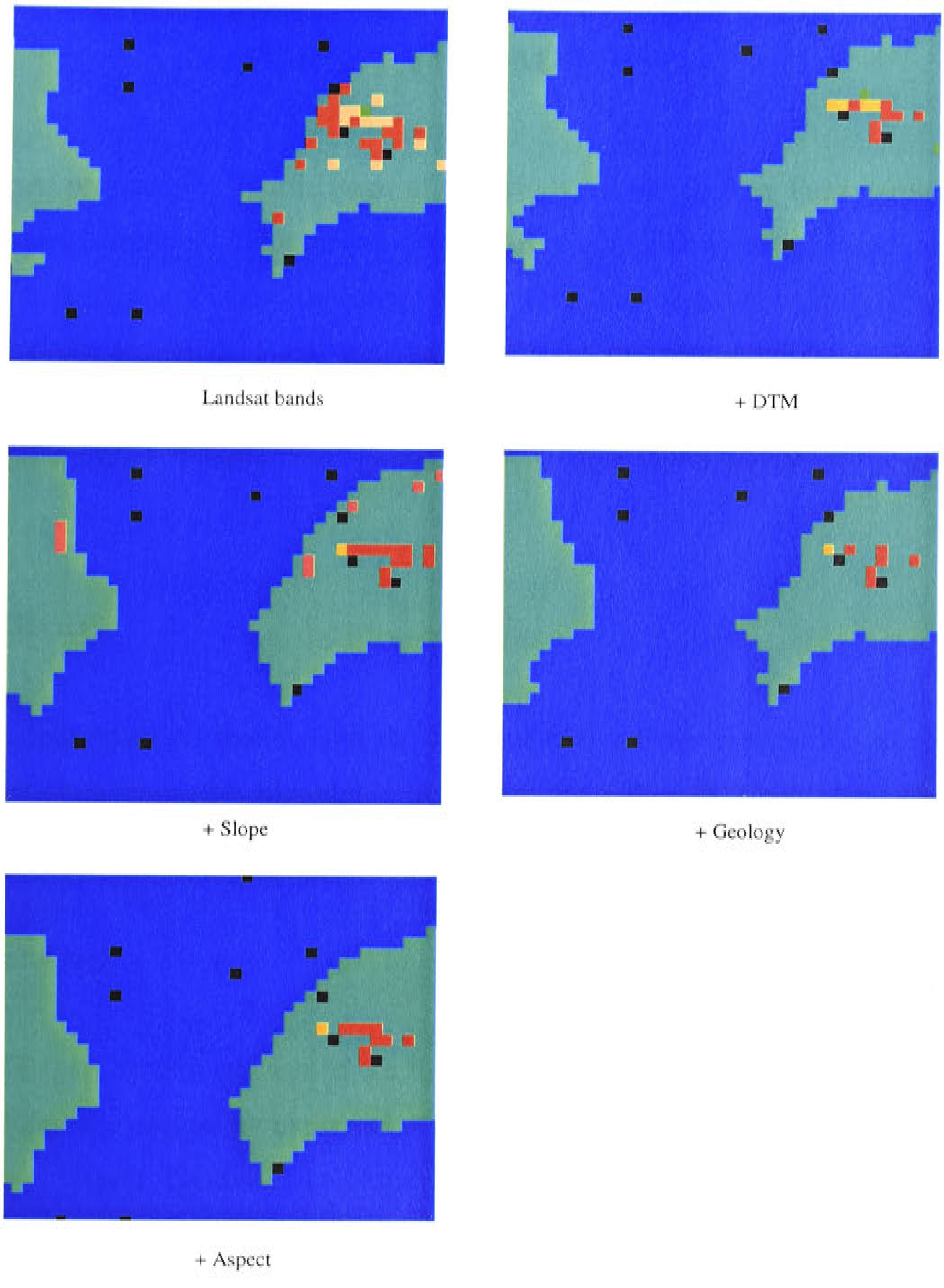


Figure 7.3 Error manifestations through the model based on Dempster-Shafer's theory by adding variables: from left to right, from top to bottom, Landsat bands, + DTM, + slope, + geology, + aspect

Chapter 8

FUZZY (RULE-BASED) EXPERT SYSTEMS FOR PREDICTIVE FOREST TYPE MAPPING

The chapter presents the fourth stage of this study, in which four Fuzzy (rule-based) Expert Systems were built from learning samples and applied to complicated predictive forest type mapping. Firstly, the chapter describes the methods to build the Fuzzy Expert Systems in detail. Then, the chapter reports the results of the Fuzzy Expert Systems for complicated forest type mapping. Following that, the chapter discusses the results and findings. Finally, the chapter gives a short summary of this stage of the study.

8.1 METHODS

The three individual AI models and the combined AI models used in this study could not present comprehensible mapping processes to the end users. Expert Systems on the other hand, can help the user interpret the mapping process by representing knowledge as classification rules and capture the reasoning behind the mapping process through using formal logic.

A typical Expert System consists of the following components: a database, a knowledge base, an inference engine, an explanation machine and a user interface (see Figure 2.2). The Fuzzy Expert Systems developed in this study contain all these essential components.

The database of the Fuzzy Expert Systems is simple. It contains input crisp data of the seven independent variables. The only difference is that they are not in GIS grid format but in the file format of FuzzyCLIPS (Orchard, 1998) that was used to program the Fuzzy Expert Systems. The database can also include fuzzified data, which was used in the Fuzzy Expert System that learns from the results of a combined AI model. Other essential components of the Fuzzy Expert Systems are described below.

8.1.1 BUILDING FUZZY KNOWLEDGE (RULE) BASES FROM SAMPLES

Traditionally, the knowledge base is derived from domain experts by knowledge engineers. The method suffers several problems. The most pronounced one is the so-called knowledge acquisition "bottleneck" problem. Alternatively, recent studies have shown that machine learning methods such as Decision Trees could be used to derive rules (e.g., Huang & Jensen, 1997). Moreover, rules can also be generated by directly learning from examples (Wang & Mendel, 1992). Generating rules from learning samples largely avoids the knowledge acquisition "bottleneck" problem. It is a relatively straightforward and quick way to build a large knowledge base.

This study adopted similar method to that of Wang and Mendel (1992) to generate fuzzy rules directly from samples. The process can be divided into the following seven steps; select appropriate learning samples, divide the input spaces into fuzzy regions, generate fuzzy rules from the learning samples, resolve potentially conflicting rules, resolve the unknown class by adding new rules and modifying the existing rules, prune rules, and code the rule bases.

8.1.1.1 Select appropriate learning samples

One way to select appropriate learning samples is to select from the results of the combined AI models. This study chose the hard classification map of D-S1 (Plate 6) as the base map because of its relatively high predictive accuracies. Pixels with high prediction probabilities and confidence values were selected from the classification map as the learning samples for each class. The quantitative confidence measure of D-S1 was based on fuzzy set theory (Plate 36). For example, for a pixel, first the differences between the probability outcomes of D-S1 and those of the three individual models were calculated and summed. Then a fuzzy membership function (Equation 8.1) was used to evaluate the confidence value of the summed differences. The smaller the summed difference, the higher the calculated confidence is.

$$\mu = \begin{cases} 1-2\times(x/3)^2 & x\in[0,1.5) \\ 2\times((3-x)/3)^2 & x\in[1.5,3) \\ 0 & elsewhere \end{cases} \qquad (8.1)$$

where μ is the fuzzy membership, and x is the summed difference

In addition, the measurements of probability or degrees of belief of the most likely class for individual pixels of D-S1 could also be displayed along with the confidence values (Plate 37).

For example, the following criterion was used to select the learning samples of forest type1; D-S1 = 1 and probds > 0.99 and confdsp > 0.95, where probds represents the layer of probability (Plate 37) and confdsp represents the layer of confidence measure (Plate 36). Table 8.1 lists the learning samples selected from D-S1. The criteria used to select the learning samples (e.g., Table 8.1) are arbitrary to a degree. The number of samples chosen is much smaller than the training samples of Table 3.1. However, the assumption was that learning samples chosen in such a way might be a good representation of each class. Therefore, the number of learning samples for each class does not have to be large.

Table 8.1 Samples selected from D-S1

Class Name	Number of samples	Selection criterion
Class 1	54	D-S1 = 1 and probds > 0.99 and confdsp > 0.95
Class 2	51	D-S1 = 2 and probds > 0.95 and confdsp > 0.88
Class 3	24	D-S1 = 3 and probds > 0.75 and confdsp > 0.775
Class 4	44	D-S1 = 4 and probds > 0.95 and confdsp > 0.815
Class 5	58	D-S1 = 5 and probds > 0.95 and confdsp > 0.92
Class 6	55	D-S1 = 6 and probds > 0.99 and confdsp > 0.942
Class 7	57	D-S1 = 7 and probds > 0.99 and confdsp > 0.97
Class 8	68	D-S1 = 8 and probds = 1.0 and confdsp > 0.998
Class 9	183	D-S1 = 9 and probds = 1.0 and confdsp > 0.99885

The other way to select appropriate learning samples is to select from the field samples. In this study, 80% of the field samples were chosen to be the learning samples of the fuzzy rule bases. The learning samples are exactly the same as the training samples of the three individual AI models to ensure the effectiveness of comparison between them (Table 3.1).

After selecting the learning samples, each of them could define an input-output pair, in which the input is the spatial space of the 7 independent variables, and the output is the class value assigned to the sample (pixel). One example of such an (input)~(output) pair is; (band2: 21, band4: 29, band7: 10, DTM: 46, slope: 4, aspect: 351, geology: 5) → (class: 1).

8.1.1.2 Divide the input spaces into fuzzy regions

In order to derive fuzzy rules from the learning samples, the initial crisp input data such as those shown in the above example need to be fuzzified into fuzzy data. This study has divided the input spaces into fuzzy regions (fuzzy linguistic values), each of which was assigned a fuzzy membership function.

The band2 data was divided into eight fuzzy regions (Figure 8.1) with fuzzy membership functions of Equations 8.2a-h, respectively.

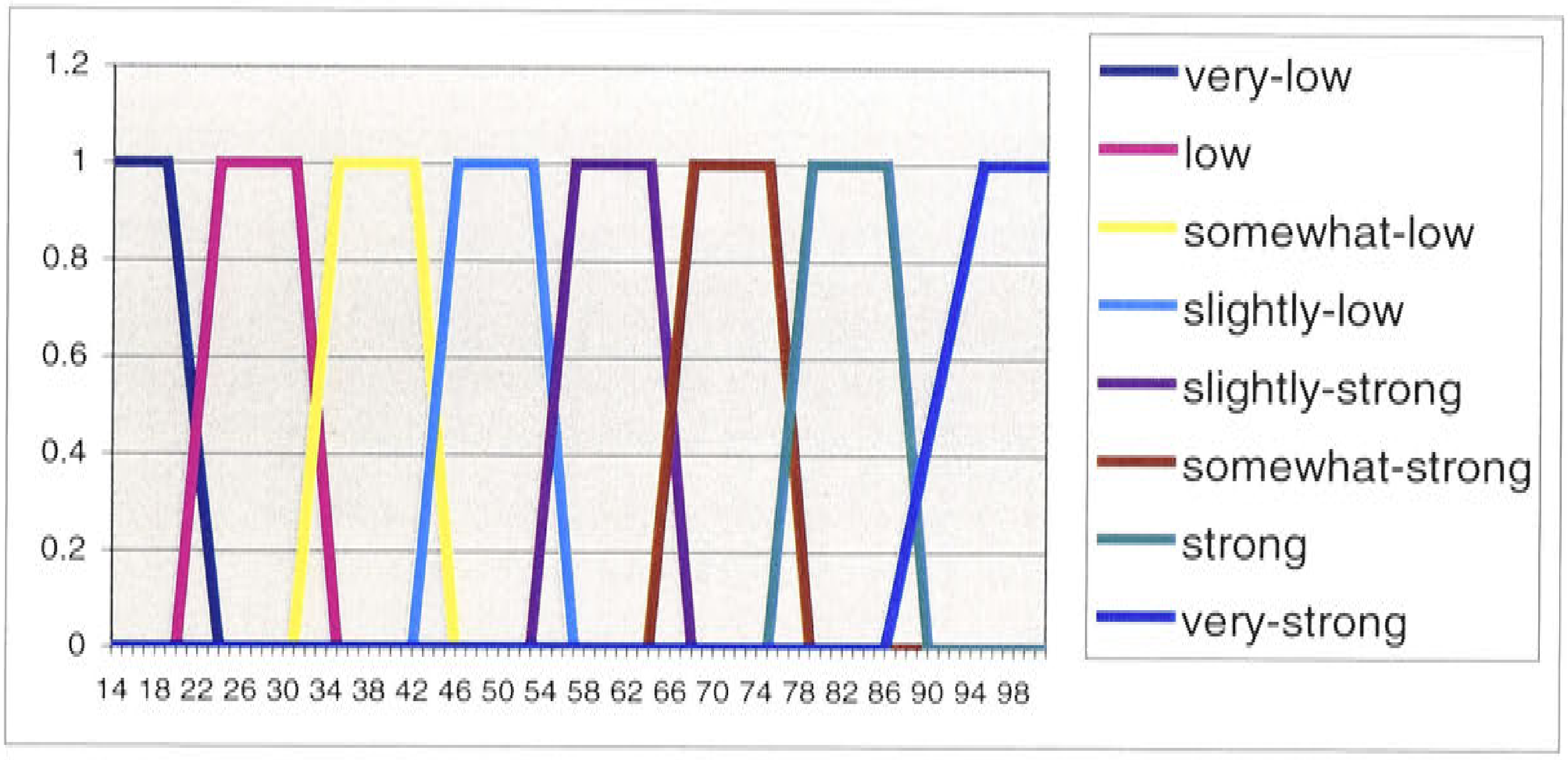


Figure 8.1 Fuzzy regions of band2

$$\mu_{band2-verylow}(x) = \begin{cases} 1 & x \in [14,19] \\ \dfrac{24-x}{5} & x \in (19,24] \\ 0 & elsewhere \end{cases} \quad (8.2a)$$

$$\mu_{band2-low}(x) = \begin{cases} \dfrac{x-20}{4} & x \in [20,24) \\ 1 & x \in [24,31) \\ \dfrac{35-x}{4} & x \in [31,35] \\ 0 & elsewhere \end{cases} \quad (8.2b)$$

$$\mu_{band2-slightlystong}(x) = \begin{cases} \dfrac{x-53}{4} & x \in [53,57) \\ 1 & x \in [57,64) \\ \dfrac{68-x}{4} & x \in [64,68] \\ 0 & elsewhere \end{cases} \quad (8.2e)$$

$$\mu_{band2-somewhatstong}(x) = \begin{cases} \dfrac{x-64}{4} & x \in [64,68) \\ 1 & x \in [68,75) \\ \dfrac{79-x}{4} & x \in [75,79] \\ 0 & elsewhere \end{cases} \quad (8.2f)$$

$$\mu_{band2-somewhatlow}(x)=\begin{cases}\frac{x-31}{4} & x\in[31,35)\\ 1 & x\in[35,42)\\ \frac{46-x}{4} & x\in[42,46]\\ 0 & elsewhere\end{cases} \quad (8.2c)$$

$$\mu_{band2-slightlylow}(x)=\begin{cases}\frac{x-42}{4} & x\in[42,46)\\ 1 & x\in[46,53)\\ \frac{57-x}{4} & x\in[53,57]\\ 0 & elsewhere\end{cases} \quad (8.2d)$$

$$\mu_{band2-strong}(x)=\begin{cases}\frac{x-75}{4} & x\in[75,79)\\ 1 & x\in[79,86)\\ \frac{86-x}{4} & x\in[86,90]\\ 0 & elsewhere\end{cases} \quad (8.2g)$$

$$\mu_{band2-verystrong}(x)=\begin{cases}\frac{x-86}{9} & x\in[86,95)\\ 1 & x\in[95,101]\\ 0 & elsewhere\end{cases} \quad (8.2h)$$

The band4 data was divided into eight fuzzy regions (Figure 8.2) with fuzzy membership functions of Equations 8.3a-h, respectively.

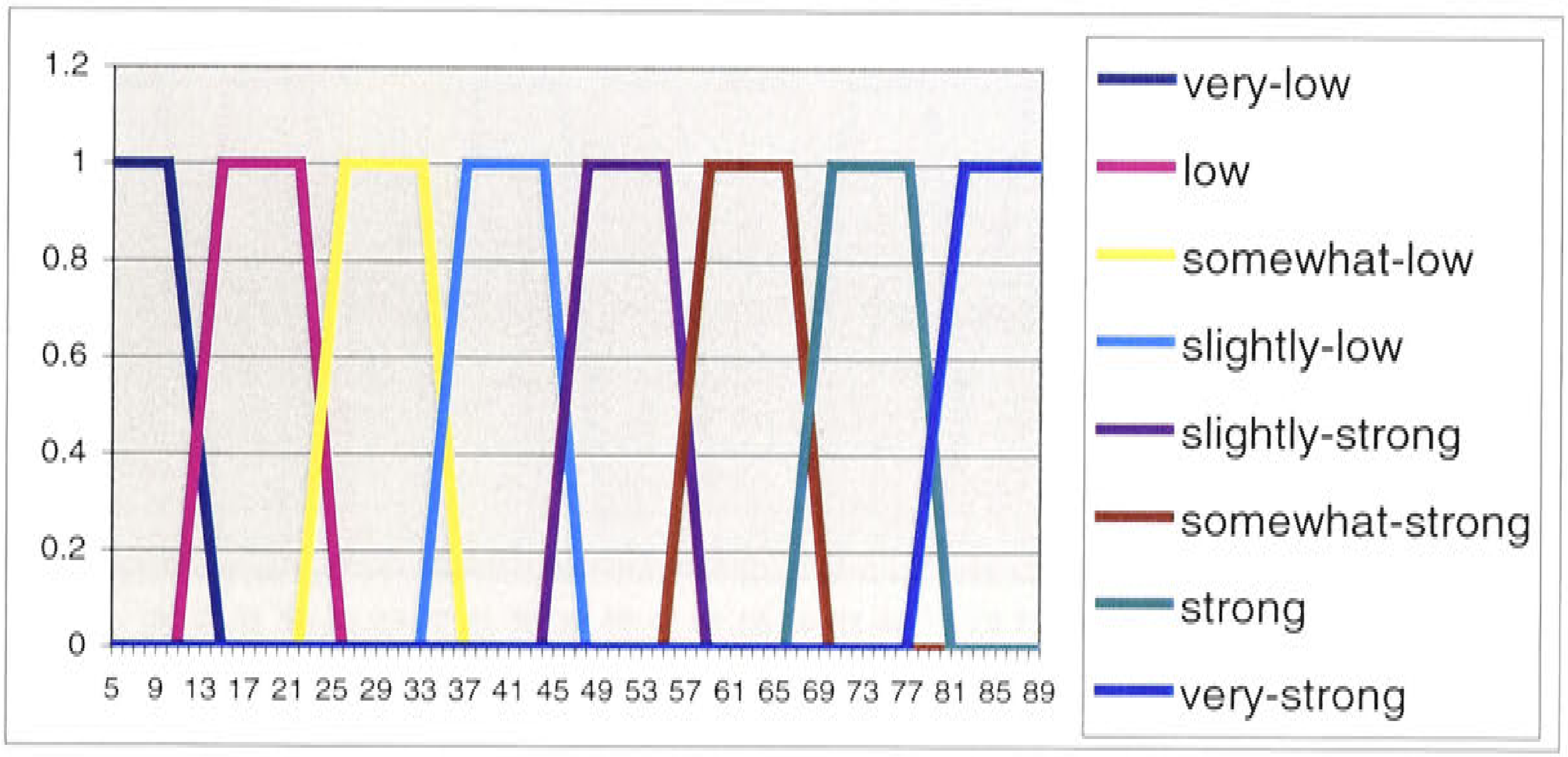


Figure 8.2 Fuzzy regions of band4

$$\mu_{band4-verylow}(x)=\begin{cases}1 & x\in[5,10]\\ \frac{15-x}{5} & x\in(10,15]\\ 0 & elsewhere\end{cases} \quad (8.3a)$$

$$\mu_{band4-low}(x)=\begin{cases}\frac{x-11}{4} & x\in[11,15)\\ 1 & x\in[15,22)\\ \frac{26-x}{4} & x\in[22,26]\\ 0 & elsewhere\end{cases} \quad (8.3b)$$

$$\mu_{band4-somewhatlow}(x)=\begin{cases}\frac{x-22}{4} & x\in[22,26)\\ 1 & x\in[26,33)\\ \frac{37-x}{4} & x\in[33,37]\\ 0 & elsewhere\end{cases} \quad (8.3c)$$

$$\mu_{band\ 4-slightlylo\ w}(x)=\begin{cases}\frac{x-33}{4} & x\in[33,37)\\ 1 & x\in[37,44)\\ \frac{48-x}{4} & x\in[44,48]\\ 0 & elsewhere\end{cases} \quad (8.3d)$$

$$\mu_{band4-slightlystrong}(x)=\begin{cases}\frac{x-44}{4} & x\in[44,48)\\ 1 & x\in[48,55)\\ \frac{59-x}{4} & x\in[55,59]\\ 0 & elsewhere\end{cases} \quad (8.3e)$$

$$\mu_{band4-somewhatstrong}(x)=\begin{cases}\frac{x-55}{4} & x\in[55,59)\\ 1 & x\in[59,66)\\ \frac{70-x}{4} & x\in[66,70]\\ 0 & elsewhere\end{cases} \quad (8.3f)$$

$$\mu_{band4-strong}(x)=\begin{cases}\frac{x-66}{4} & x\in[66,70)\\ 1 & x\in[70,77)\\ \frac{81-x}{4} & x\in[77,81]\\ 0 & elsewhere\end{cases} \quad (8.3g)$$

$$\mu_{band4-verystrong}(x)=\begin{cases}\frac{x-77}{5} & x\in[77,82)\\ 1 & x\in[82,89]\\ 0 & elsewhere\end{cases} \quad (8.3h)$$

The band7 data was divided into eight fuzzy regions (Figure 8.3) with fuzzy membership functions of Equations 8.4a-h, respectively.

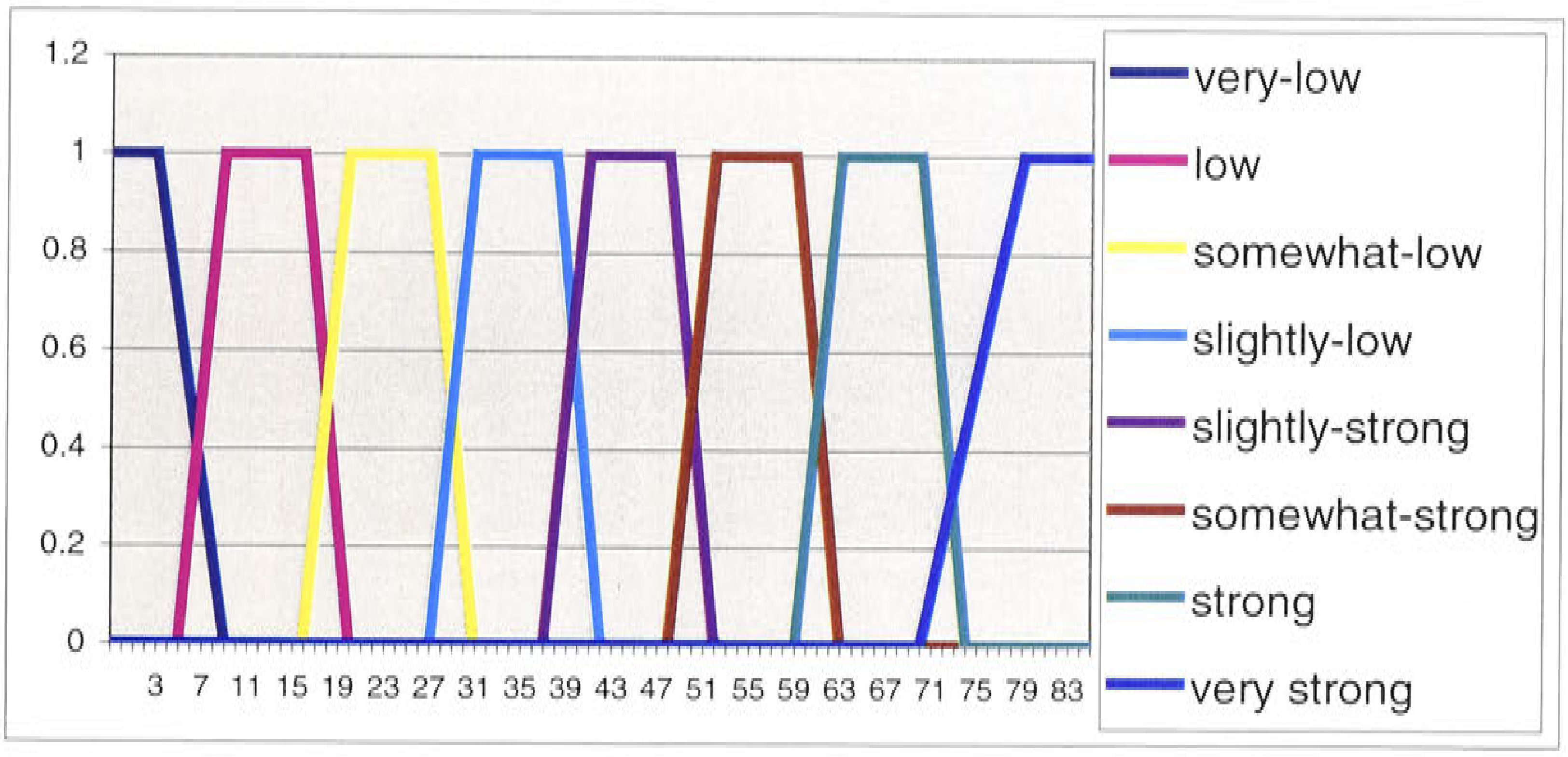


Figure 8.3 Fuzzy regions of band7

$$\mu_{band7-verylow}(x)=\begin{cases}1 & x\in[1,4]\\ \frac{10-x}{6} & x\in(4,10]\\ 0 & elsewhere\end{cases} \quad (8.4a)$$

$$\mu_{band7-low}(x)=\begin{cases}\frac{x-6}{4} & x\in[6,10)\\ 1 & x\in[10,17)\\ \frac{21-x}{4} & x\in[17,21]\\ 0 & elsewhere\end{cases} \quad (8.4b)$$

$$\mu_{band7-somewhatlow}(x)=\begin{cases}\frac{x-17}{4} & x\in[17,21)\\ 1 & x\in[21,28)\\ \frac{32-x}{4} & x\in[28,32]\\ 0 & elsewhere\end{cases} \quad (8.4c)$$

$$\mu_{band7-slightlylow}(x)=\begin{cases}\frac{x-28}{4} & x\in[28,32)\\ 1 & x\in[32,39)\\ \frac{43-x}{4} & x\in[39,43]\\ 0 & elsewhere\end{cases} \quad (8.4d)$$

$$\mu_{band7-slightlystong}(x)=\begin{cases}\frac{x-38}{4} & x\in[38,42)\\ 1 & x\in[42,49)\\ \frac{53-x}{4} & x\in[49,53]\\ 0 & elsewhere\end{cases} \quad (8.4e)$$

$$\mu_{band7-somewhatstrong}(x)=\begin{cases}\frac{x-49}{4} & x\in[49,53)\\ 1 & x\in[53,60)\\ \frac{64-x}{4} & x\in[60,64]\\ 0 & elsewhere\end{cases} \quad (8.4f)$$

$$\mu_{band7-strong}(x)=\begin{cases}\frac{x-60}{4} & x\in[60,64)\\ 1 & x\in[64,71)\\ \frac{75-x}{4} & x\in[71,75]\\ 0 & elsewhere\end{cases} \quad (8.4g)$$

$$\mu_{band7-verystrong}(x)=\begin{cases}\frac{x-71}{9} & x\in[71,80)\\ 1 & x\in[80,86]\\ 0 & elsewhere\end{cases} \quad (8.4h)$$

The DTM data was divided into nine fuzzy regions (Figure 8.4) with fuzzy membership functions of Equations 8.5a-i, respectively.

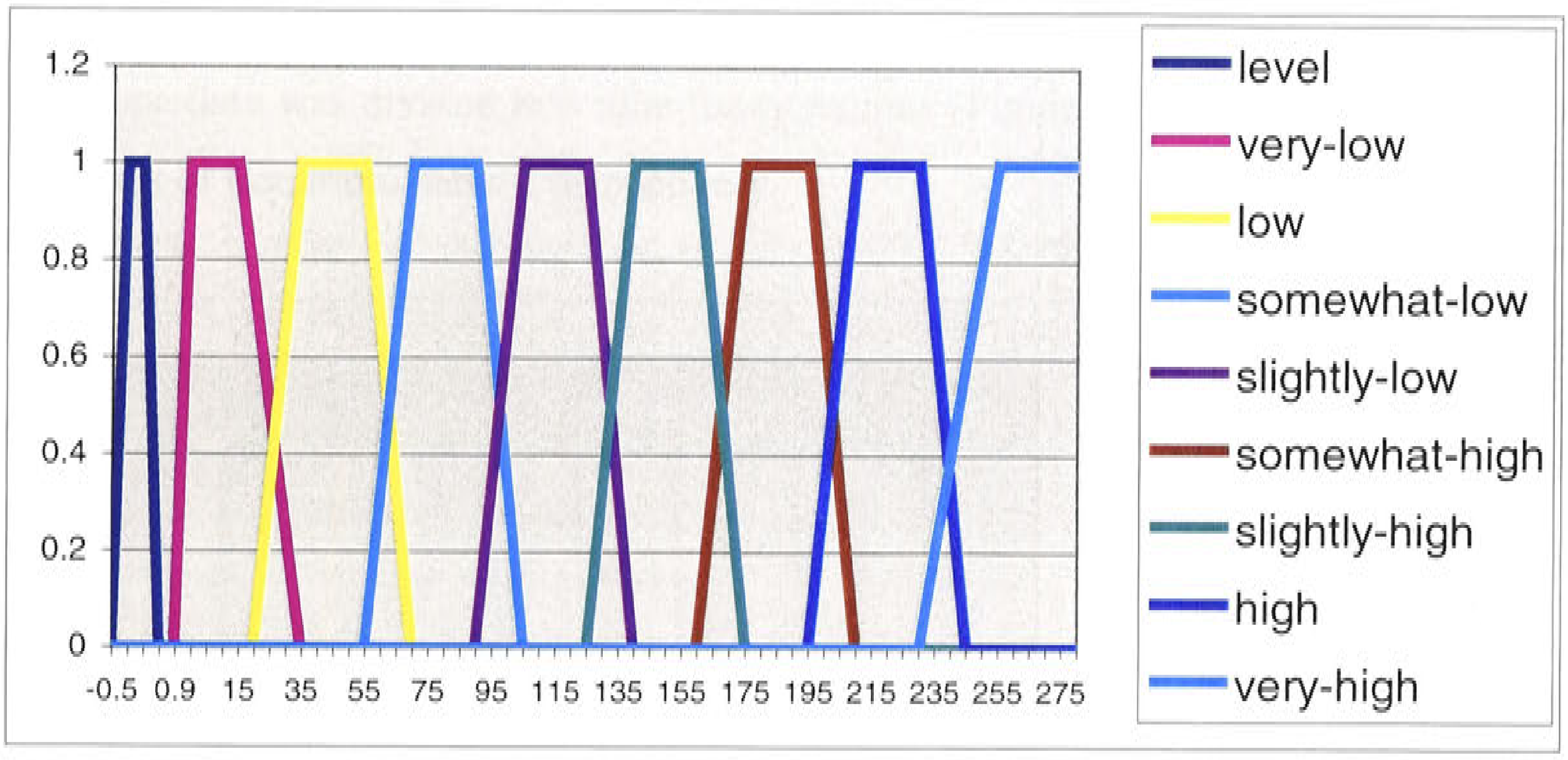


Figure 8.4 Fuzzy regions of DTM

$$\mu_{dtm-level}(x)=\begin{cases}\frac{x+0.5}{0.4} & x\in[-0.5,-0.1)\\ 1 & x\in[-0.1,0.1)\\ \frac{0.5-x}{0.4} & x\in[0.1,0.5)\\ 0 & elsewhere\end{cases} \quad (8.5a)$$

$$\mu_{dtm-verylow}(x)=\begin{cases}\frac{x-0.9}{0.1} & x\in(0.9,1]\\ 1 & x\in(1,15]\\ \frac{35-x}{20} & x\in[15,35)\\ 0 & elsewhere\end{cases} \quad (8.5b)$$

$$\mu_{dtm-low}(x)=\begin{cases}\frac{x-20}{15} & x\in[20,35)\\ 1 & x\in[35,55)\\ \frac{70-x}{15} & x\in[55,70]\\ 0 & elsewhere\end{cases} \quad (8.5c)$$

$$\mu_{dtm-somewhatlow}(x)=\begin{cases}\frac{x-55}{15} & x\in[55,70)\\ 1 & x\in[70,90)\\ \frac{105-x}{15} & x\in[90,105]\\ 0 & elsewhere\end{cases} \quad (8.5d)$$

$$\mu_{dtm-slightlylow}(x) = \begin{cases} \frac{x-90}{15} & x \in [90,105) \\ 1 & x \in [105,125) \\ \frac{140-x}{15} & x \in [125,140] \\ 0 & elsewhere \end{cases} \quad (8.5e)$$

$$\mu_{dtm-slightlyhigh}(x) = \begin{cases} \frac{x-125}{15} & x \in [125,140) \\ 1 & x \in [140,160) \\ \frac{175-x}{15} & x \in [160,175] \\ 0 & elsewhere \end{cases} \quad (8.5f)$$

$$\mu_{dtm-somewhathigh}(x) = \begin{cases} \frac{x-160}{15} & x \in [160,175) \\ 1 & x \in [175,195) \\ \frac{210-x}{15} & x \in [195,210] \\ 0 & elsewhere \end{cases} \quad (8.5g)$$

$$\mu_{dtm-high}(x) = \begin{cases} \frac{x-195}{15} & x \in [195,210) \\ 1 & x \in [210,230) \\ \frac{245-x}{15} & x \in [230,245] \\ 0 & elsewhere \end{cases} \quad (8.5h)$$

$$\mu_{dtm-veryhigh}(x) = \begin{cases} \frac{x-230}{25} & x \in [230,255) \\ 1 & x \in [255,280] \\ 0 & elsewhere \end{cases} \quad (8.5i)$$

The slope data was divided into nine fuzzy regions (Figure 8.5) with fuzzy membership functions of Equations 8.6a-i, respectively.

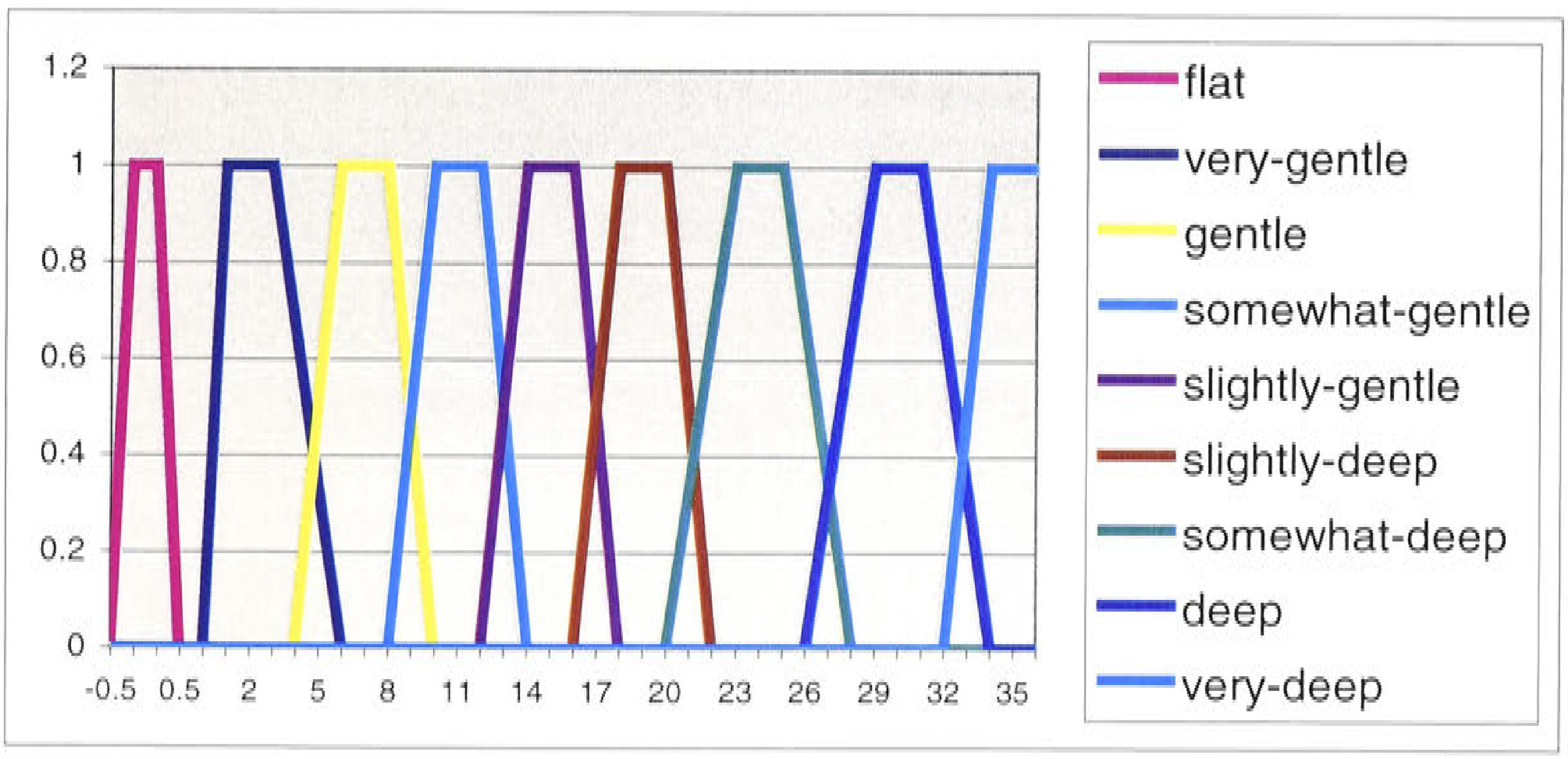


Figure 8.5 Fuzzy regions of slope

$$\mu_{slope-flat}(x)=\begin{cases}\dfrac{x+0.5}{0.4} & x\in[-0.5,-0.1)\\ 1 & x\in[-0.1,0.1)\\ \dfrac{0.5-x}{0.4} & x\in[0.1,0.5)\\ 0 & elsewhere\end{cases} \quad (8.6a)$$

$$\mu_{slope-verygentle}(x)=\begin{cases}\dfrac{x-0.9}{0.1} & x\in(0.9,1]\\ 1 & x\in(1,3]\\ \dfrac{6-x}{3} & x\in[3,6)\\ 0 & elsewhere\end{cases} \quad (8.6b)$$

$$\mu_{slope-gentle}(x)=\begin{cases}\dfrac{x-4}{2} & x\in[4,6)\\ 1 & x\in[6,8)\\ \dfrac{10-x}{2} & x\in[8,10]\\ 0 & elsewhere\end{cases} \quad (8.6c)$$

$$\mu_{slope-somewhatgentle}(x)=\begin{cases}\dfrac{x-8}{2} & x\in[8,10)\\ 1 & x\in[10,12)\\ \dfrac{14-x}{2} & x\in[12,14]\\ 0 & elsewhere\end{cases} \quad (8.6d)$$

$$\mu_{slope-slightlygentle}(x)=\begin{cases}\dfrac{x-12}{2} & x\in[12,14)\\ 1 & x\in[14,16)\\ \dfrac{18-x}{2} & x\in[16,18]\\ 0 & elsewhere\end{cases} \quad (8.6e)$$

$$\mu_{slope-slightlydaep}(x)=\begin{cases}\dfrac{x-16}{2} & x\in[16,18)\\ 1 & x\in[18,20)\\ \dfrac{22-x}{2} & x\in[20,22]\\ 0 & elsewhere\end{cases} \quad (8.6f)$$

$$\mu_{slope-somewhatde\,ep}(x)=\begin{cases}\dfrac{x-20}{3} & x\in[20,23)\\ 1 & x\in[23,25)\\ \dfrac{28-x}{3} & x\in[25,28]\\ 0 & elsewhere\end{cases} \quad (8.6g)$$

$$\mu_{slope-deep}(x)=\begin{cases}\dfrac{x-26}{3} & x\in[26,29)\\ 1 & x\in[29,31)\\ \dfrac{34-x}{3} & x\in[31,34]\\ 0 & elsewhere\end{cases} \quad (8.6h)$$

$$\mu_{slope-verydeep}(x)=\begin{cases}\dfrac{x-32}{2} & x\in[32,34)\\ 1 & x\in[34,36]\\ 0 & elsewhere\end{cases} \quad (8.6i)$$

The aspect data was divided into nine fuzzy regions (Figure 8.6) with fuzzy membership functions of Equations 8.7a-i, respectively.

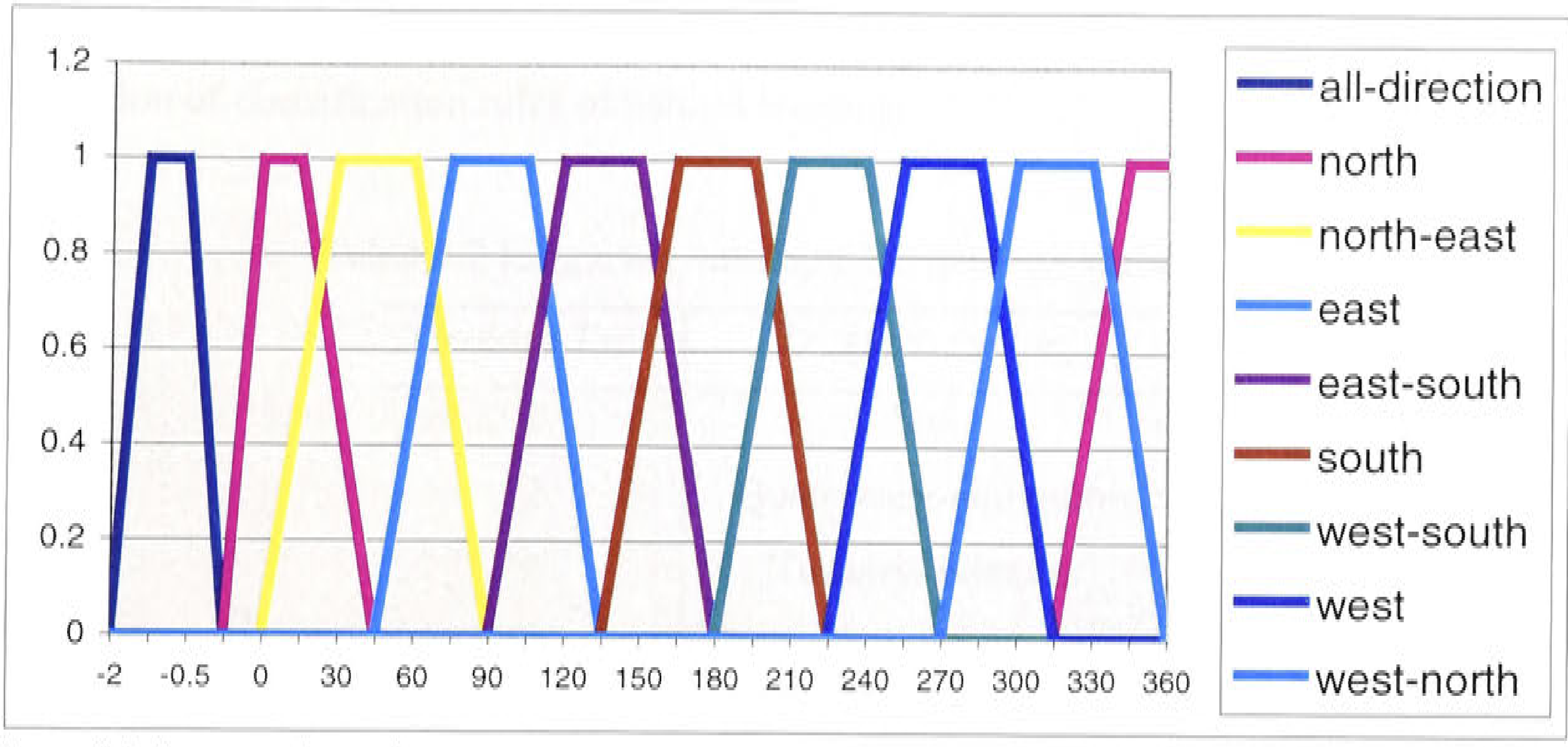


Figure 8.6 Fuzzy regions of aspect

$$\mu_{aspect-all-direction}(x)=\begin{cases}\frac{x+2}{0.5} & x\in[-2,-1.5)\\ 1 & x\in[-1.5,-0.5)\\ \frac{-0.1-x}{0.4} & x\in[-0.5,-0.1)\\ 0 & elsewhere\end{cases} \quad (8.7a)$$

$$\mu_{aspect-north}(x)=\begin{cases}\frac{x+0.1}{0.1} & x\in[-0.1,0)\\ 1 & x\in[0,15]\\ \frac{45-x}{30} & x\in(15,45]\\ \frac{x-315}{30} & x\in[315,345)\\ 1 & x\in[345,360)\\ 0 & elsewhere\end{cases} \quad (8.7b)$$

$$\mu_{aspect-north-east}(x)=\begin{cases}\frac{x}{30} & x\in[0,30)\\ 1 & x\in[30,60)\\ \frac{90-x}{30} & x\in[60,90]\\ 0 & elsewhere\end{cases} \quad (8.7c)$$

$$\mu_{aspect-east}(x)=\begin{cases}\frac{x-45}{30} & x\in[45,75)\\ 1 & x\in[75,105)\\ \frac{135-x}{30} & x\in[105,135]\\ 0 & elsewhere\end{cases} \quad (8.7d)$$

$$\mu_{aspect-east-south}(x)=\begin{cases}\frac{x-90}{30} & x\in[90,120)\\ 1 & x\in[120,150)\\ \frac{180-x}{30} & x\in[150,180]\\ 0 & elsewhere\end{cases} \quad (8.7e)$$

$$\mu_{aspect-south}(x)=\begin{cases}\frac{x-135}{30} & x\in[135,165)\\ 1 & x\in[165,195)\\ \frac{225-x}{30} & x\in[195,225]\\ 0 & elsewhere\end{cases} \quad (8.7f)$$

$$\mu_{aspect-west-south}(x)=\begin{cases}\frac{x-180}{30} & x\in[180,210)\\ 1 & x\in[210,240)\\ \frac{270-x}{30} & x\in[240,270]\\ 0 & elsewhere\end{cases} \quad (8.7g)$$

$$\mu_{aspect-west}(x)=\begin{cases}\frac{x-225}{30} & x\in[225,255)\\ 1 & x\in[255,285)\\ \frac{315-x}{30} & x\in[285,315]\\ 0 & elsewhere\end{cases} \quad (8.7h)$$

$$\mu_{aspect-west-north}(x)=\begin{cases}\frac{x-270}{30} & x\in[270,300)\\ 1 & x\in[300,330)\\ \frac{360-x}{30} & x\in[330,360)\\ 0 & elsewhere\end{cases} \quad (8.7i)$$

The geology variable is categorical data that does not need to be fuzzified. Nevertheless, its initial data values have been transformed into linguistic values (Table 8.2) for the generation of classification rules of natural language.

Table 8.2 Linguistic values of the geology variable

Geology Type	*Linguistic value*
1	Sea
2	Quaternary-alluvium
3	Tertiary-essexite
4	Snapper-point-permian
5	Pebbly-beach-permian
6	Wasp-head-permian
7	Ordovician

8.1.1.3 Generate fuzzy rules from the learning samples

After defining the fuzzy regions, the degrees of any input crisp data belonging to different fuzzy regions were calculated. For example, for "band7 = 8", it falls in two fuzzy regions; 'verylow' and 'low' with fuzzy membership (degrees of belongings) of 0.3333 (i.e., $\mu_{band7\text{-}verylow}(8) = 0.3333$) and 0.5 (i.e., $\mu_{band7\text{-}low}(8) = 0.5$) respectively. Then, the fuzzy region or linguistic value with maximum fuzzy membership was assigned to the input data. In the above example, the result is; band7(8) = low.

Consequently, for each of the learning sample, it created a (fuzzy input)~(crisp output) pair. For example, the (input)~(output) pair illustrated in section 4.4.1.1 has been converted to a (fuzzy input)~(crisp output) pair as; (band2: very-low, band4: somewhat-low, band7: low, DTM: low, slope: very-gentle, aspect: north, geology: pebbly-beach-permian) → (class: 1). This fuzzy input–crisp output pair could be easily turned into a fuzzy rule such as:

If band2 is very-low

and band4 is somewhat-low

and band7 is low

and DTM is low

and slope is very-gentle

and aspect is north

and geology is pebbly-beach-permian

Then the class is 1

The rule has 7 fuzzy antecedents and a single crisp consequence that is called; fuzzy-crisp rule.

Using this method, this study has generated fuzzy rules from the learning samples, each of which corresponds to a sample. However, because many of the fuzzy rules generated in such a way were duplicated, they have been deleted, but the number of learning samples each rule was generated from was counted.

8.1.1.4 Resolve potentially conflicting rules

Because of the large learning samples and possible data errors, the fuzzy rules generated in such a way are potentially in conflict, as some of these rules have the same antecedents but different consequences. Resolving conflicts is difficult and time consuming.

In this study, the author decided that usually only one linguistic value could be assigned to the antecedents of band2, band4, band7, DTM and slope. But for the antecedents of aspect and geology more than one linguistic value could be assigned by using *or* operator to facilitate the resolving of conflicts. Therefore, when the fuzzy rules have the same linguistic value of band2, band4, band7, DTM and slope in their antecedents, they were grouped and marked as potentially conflicting rules. Then, the following factors were considered to resolve the potential conflicts with decreased priorities:

- The fuzzy linguistic values of aspect and geology,
- The proportion of learning samples the rule was generated from, which is not the absolute number of samples recorded but rather the percentage of the number out of the total learning samples of the class, and
- The confidence of the rule, which equals to the minimum fuzzy membership of the linguistic values.

For example, if two rules marked as potentially conflicting rules have different linguistic values of aspect or geology they are not conflicting any way. But if they do have the same linguistic values of aspect and geology then the rule generated from higher proportion of samples take the priority. However, if they are generated from a similar proportion of samples then the rule with higher confidence take the priority.

Having resolved the potential conflicts, those rules marked as potentially conflicting rules would have the same linguistic values of band2, band4, band7, DTM and slope, and different (maybe multiple) linguistic values of aspect and/or geology in their antecedents. Up to this step, two fuzzy rule bases have been built. One was generated from the learning samples selected from the combined AI model of D-S1, which is named **fes12**. The other was generated from the learning samples that selected from the field samples, which is named **sarules**.

8.1.1.5 Resolve the unknown class by adding new rules and modifying the existing rules

It should be noted that both of the rule bases (fes12 and sarules) contain a rule to accommodate for an unknown class. This is necessary because the input data at the sites of the learning samples do not cover all possible input values, which in turn has generated incomplete fuzzy rule bases.

Applying the two rule bases to the 20% test data did show that there were inputs yielding the unknown class (class 10 in this case). These cases of unknown class could be resolved to improve the completeness of the rule bases. This study resolved the cases of the unknown class by adding new rules and modifying the existing rules. These rules were generated and modified from the test samples that were predicted as the unknown class. This is done by using the same three steps described above; divide the input spaces into fuzzy regions, generate fuzzy rules from samples, and resolve potential conflicting rules. After resolving the unknown class by adding new rules and modifying the existing rules, a new fuzzy rule base was generated and named **sarules4**. The above procedures did not apply to fes12, because the test accuracy of fes12 was so low that the author believes that it is not possible to make large improvements (see Table 5.20).

8.1.1.6 Prune rules from sarules4

The fuzzy rule base of sarules4 built in the last step is large and potentially over-specified. Therefore, pruning is necessary to increase the generalization of the rule base, similar to the process often used in the applications of Decision Trees and Artificial Neural Networks. In this study, pruning was used to cut rules that are less significant or were generated from few learning samples.

The following two criteria were used to prune rules from sarules4:

- If a fuzzy rule was not marked as a potentially conflicting rule and it was generated from only one learning sample, then it was removed from the rule base except that it is a class2 rule or a class3 rule.
- If a group of fuzzy rules were marked as a potentially conflicting rules, then remove rules which satisfying the following conditions from the group:
    1. It is a class1 rule generated from less than four learning samples.
    2. It is a class2 rule generated from only one learning sample.
    3. It is a class3 rule generated from only one learning sample.
    4. It is a class4 rule generated from less than three learning samples.
    5. It is a class5 rule generated from less than three learning samples.
    6. It is a class6 rule generated from only one learning sample.
    7. It is a class7 rule generated from only one learning sample.
    8. It is a class8 rule generated from less than three learning samples.

  Except if the group is supposed to be removed entirely from the rule base by applying the above conditions, and there are class2 rule or/and class3 rule in the original group, then the class2 rule and class3 rule are left in the rule base even they were generated from only one learning sample.

The reason to preserve the class2 and class3 rules is that there are fewer learning samples for these two classes, which make them potentially biased to the classes with large samples.

After pruning a large number of rules from sarules4, it would obviously cause a large unclassified area (e.g., area assigned to the unknown class). To reduce the unclassified area, several rules have been added to the rule base. The resultant rule base was named **sarules7**.

8.1.1.7 Code the rule bases

The above rule bases were turned into computer code using program language of FuzzyCLIPS (Orchard, 1998) which is an Expert System language developed by the Artificial Intelligence Section, Lyndon B.Johnson Space Center, NASA.

An example of a fuzzy rule after coding is as follows:

```
(defrule class7rule47
  (band2 (band2 very-low) (location ?i))
  (band4 (band4 low) (location ?i))
  (band7 (band7 very-low) (location ?i))
  (elevation (DTM slightly-low) (location ?i))
  (gradient (slope somewhat-gentle) (location ?i))
  (or (aspect (aspect west) (location ?i))
      (aspect (aspect north) (location ?i)))
  (geology (geology quatenary-alluvium | tertiary-essexite | snapper-point-permian | pebbly-
            beach-permian | ordovician) (location ?i))
  =>
 (assert (class (value 7) (name Rainforest) (location ?i))))
```

where "defrule" defines a rule construct, "class7rule47" is the rule name, the "=>" symbol means inferring (or *Then*), the fuzzy antecedents of the rule are set before the "=>" symbol, and the crisp consequence which asserts a class value and name is displayed after the "=>" symbol.

8.1.2 INFERENCE ENGINE OF THE FUZZY EXPERT SYSTEMS

The inference engine of a Fuzzy (rule-based) Expert System is responsible for handling partial matching, deducing conclusions, calculating certainty factor of the conclusions, and resolving conflicting results. Under the circumstance of fuzzy rule bases, approximate reasoning instead of exact reasoning becomes the only appropriate inference method (see section 2.5.3). In this study, fuzzy logic or the rule of generalized *modus ponents* (Zadeh, 1983) was used for approximate reasoning, as it is employed in FuzzyCLIPS.

FuzzyCLIPS uses forward chaining for its inference engine, which reasons from facts to conclusions. Because of the fuzzy antecedents of the rules, partial matching between the fuzzified input facts and the fuzzy antecedents is always encountered. FuzzyCLIPS handles partial matching by measuring the similarity between the fuzzified input facts and the fuzzy antecedents according to the following formula:

$$S = \begin{cases} P(F_\alpha \mid F_\alpha') & \text{if } N(F_\alpha \mid F_\alpha') > 0.5 \\ (N(F_\alpha \mid F_\alpha') + 0.5) \times P(F_\alpha \mid F_\alpha') & \text{otherwise} \end{cases} \tag{8.8}$$

where S is the measure of similarity, P is the measure of possibility, N is the measure of necessity, F_α is the fuzzy set of a fuzzy antecedent, F_α' is the fuzzy set of a fuzzified input fact, and

$$P(F_\alpha \mid F_\alpha') = \max(\min(\mu_{F_\alpha}(x), \mu_{F_\alpha}'(x))) \quad \forall x \in U \tag{8.9}$$

$$N(F_\alpha \mid F_\alpha') = 1 - P(\overline{F}_\alpha \mid F_\alpha') \tag{8.10}$$

where $\mu_{F_\alpha}(x)$ and $\mu_{F_\alpha}'(x)$ are the fuzzy membership functions of F_α and F_α', U is the universal set, $\overline{F}_\alpha$ is the complement of F_α described by the following membership function:

$$\mu_{\overline{F}_\alpha}(x) = 1 - \mu_{F_\alpha}(x) \quad \forall x \in U \tag{8.11}$$

Meanwhile, FuzzyCLIPS uses the MIN operation to calculate the certainty factor of the deduced conclusion. The following formula is used when involving multiple antecedents:

$$CF_c = CF_r \times \min_{i=1}^{n}(CF_{f_i} \times S_i) \tag{8.12}$$

where CF_c is the certainty factor of the deduced conclusion, CF_r is the certainty factor assigned to the rule, CF_{f_i} is the certainty factor associated with the i_{th} fuzzified input fact, S_i is the measure of similarity between the i_{th} fuzzified input fact and the associated fuzzy antecedent in the rule, n is the number of input variables (fuzzy antecedents).

When there is more than one rule fired from the same set of input data, which results in conflicting conclusions, the conclusion with the maximum CF_c is chosen as the final conclusion, and the CF_c could be recorded. Because these rules assert a single crisp consequence, no defuzzification process is needed.

8.1.3 EXPLANATION MACHINE AND USER INTERFACE OF THE FUZZY EXPERT SYSTEMS

One important advantage of an Expert System over other AI models is that an Expert System can explain its reasoning process in a simple and easily understandable way. Therefore, an explanation machine is a very important component of an Expert System which can respond to users questions such as "why do you need these information input?", "how is the conclusion obtained?", and "what if I change the fact to...?". Meanwhile, a user interface is used to facilitate communication between users and an Expert System using English like natural language, which makes the Expert System become user-friendly.

In this study, FuzzyCLIPS was used to program the user interface, and the explanation machine is realised in the user interface. The user interface is driven by command lines, and it employs a "question-response" mode to communicate with users. The user interface was not intended to handle all kinds of inquires from the user, but it contains sufficient usage indications to facilitate basic queries that are often encountered. A basic knowledge of FuzzyCLIPS is not needed. However, one requirement for the user interface to work properly is that the user needs to follow the usage indications correctly each time.

The user interface contains the following features:

1. The user interface accepts data input from both keyboard and data file.
2. The user interface accepts both fuzzy input and crisp input. The mixed fuzzy data and crisp data can be entered during a single run. For example, the user can first enter crisp data of 20 for band2 and then enter fuzzy data of "low with CF = 0.8" for band4, and so on.
3. The user interface can deal with incomplete input data in a proper way. Therefore, under the circumstance that some input data is missing or unavailable, the system is not hindered and still could generate possible answers.
4. The user interface can answer questions like "why do you need the information input?" by explaining the usefulness of the particular input variable.
5. The user interface can list the rules that partially match the input facts and the certainty factors that are associated with the conclusions of these rules.

6. The user interface can display the rule that deduces the final conclusion for the purpose of explaining the inference process behind the conclusion. This is designed to answer questions like "how the conclusion is obtained?".
7. The user interface can handle user queries like "what if I change the fact to...?" by asking the user to enter new facts then deduce new conclusion from the new input. The query process can continue until the user has chosen to quit.
8. The user interface enables the system to learn new rules or to update the rule base from expert users. For example, in the user interface, if a conclusion is judged to be right by an expert user, then the user can choose to continue the query or to quit. While, if a conclusion is judged to be wrong by an expert user, then the user interface will ask the expert user to provide the right answer. By doing this, a new rule is learned and the rule base is enlarged and updated. This is an important feature of the user interface, because good AI systems including Expert Systems are expected to be able to learn from examples and errors.

8.2 RESULTS

8.2.1 Classification results

Table 8.3 lists the number of rules in each of the fuzzy rule bases created from the two sets of learning samples.

Table 8.3 Number of rules in each of fuzzy rule bases

	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Total
Fes12	9	1	11	21	14	6	26	30	15	133
Sarules	42	16	15	59	50	32	31	43	23	311
Sarules4	42	16	16	62	51	33	32	45	25	322
Sarules7	23	14	10	37	28	16	17	23	19	187

The rule base of fes12 was generated from 594 learning samples (see Table 8.1), which means that on average each rule was generated from about 4.5 samples. The rule base of sarules was generated from 1361 learning samples, which results in about 4.4 samples per rule. However, after pruning, the size of the rule base of sarules7 was significantly reduced, and it contains only 187 rules. In other words, each of its rules was generated from average about 7.3 learning samples. These results indicate that the Fuzzy Expert Systems of fes12, sarules and sarules4 may be over-specified.

Like the other AI models mentioned above, the four Fuzzy Expert Systems have been applied to forest type mapping. Due to the very large database and the potentially large number of partial matchings for each pixel (The study area has 275625 pixels in total), the Fuzzy Expert Systems could not be applied in a single run in my machine (520 MB in memory, 1.9 GHZ CPU). Therefore, the whole study area was divided into either 5 or 10 subsets and run separately, and the results of each subset were then merged together. The following table (Table 8.4) reports the number of subsets, the total number of rules fired, and the sum of the CPU time of each Fuzzy Expert System.

Table 8.4 Lists of the number of subsets, the total number of rules fired, and the sum of the CPU time of each Fuzzy Expert System

	Number of subsets	Total number of rules fired	Sum of CPU time (h)
Fes12	5	3,320,153	88:45:15
Sarules	10	2,319,983	66:30:28
Sarules4	10	2,330,022	67:41:04
Sarules7	10	2,293,150	62:05:58

The table indicates that the most important factor that affects the CPU time is the number of subsets. For example, when the whole study area was divided into 10 instead of 5 subsets, the total number of rules fired and the sum of CPU time were significantly reduced. In addition, the size of fuzzy rule base also has slightly positive relationship with the sum of CPU time. Smaller fuzzy rule base of sarules7 used less CPU time to finish the application than sarules4 and sarules.

The error matrix of fes12 is shown in Table 8.5, which shows a very unsuccessful prediction. The Fuzzy Expert System of sarules, however, achieved an overall accuracy of 41.2% which is the same as that of the Artificial Neural Network and a Kappa accuracy of 33.3% that is slightly better than that of the Artificial Neural Network (Table 8.6). But, there were relatively large numbers of test samples remaining unclassified. After resolving the unclassified samples or the unknown classes, the Fuzzy Expert System of sarules4 increased overall accuracy to 47.4% and Kappa accuracy to 40.1%, both of which are better than those of the Decision Tree (Table 8.7). It should be noted that the remaining 2 unclassified samples in Table 8.7 are part of the 4 error samples identified in section 3.2. Therefore, it is more reasonable for them to be classified as the unknown class than the Water/Sea class. Table 8.8 reports the error matrix of sarules7 that resulted from pruning the rule base of sarules4. It indicates that after pruning a large number of rules, overall accuracy decreased about 4.4%, and Kappa accuracy decreased over 4.8%. In addition, the unclassified test samples also increased.

Table 8.5 Error matrix of fes12

		Reference Data										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Total**	**User's accuracy**
Classified Data	**1**	**9**	2	1	6	4	0	6	1	0	29	0.31
	2	0	**6**	0	0	0	0	0	0	0	6	1.00
	3	18	5	**4**	2	3	1	1	3	0	37	0.11
	4	0	0	0	**7**	6	6	1	0	0	20	0.35
	5	18	2	4	25	**27**	6	3	0	0	85	0.32
	6	8	1	3	4	2	**4**	4	0	0	26	0.15
	7	0	0	0	1	0	0	**2**	0	0	3	0.67
	8	9	5	2	1	1	3	0	21	0	42	0.50
	9	0	0	0	0	0	0	0	0	86	86	1.00
	10	0	0	1	1	0	0	3	5	3	13	
	Total	62	21	15	47	43	20	20	30	89	347	
	Producer's accuracy	0.15	0.29	0.27	0.15	0.63	0.20	0.10	0.70	0.97		
	Overall accuracy for 7 forest types	0.258772										
	Kappa accuracy for 7 forest types	0.179382										
	Kappa variance for 7 forest types	0.000909										

Table 8.6 Error matrix of sarules

	Reference Data										
Classified Data	1	2	3	4	5	6	7	8	9	Total	User's accuracy
1	**39**	4	5	13	11	1	1	0	0	74	0.53
2	4	**5**	2	0	1	0	2	0	0	14	0.36
3	3	4	**2**	1	3	2	0	0	0	15	0.13
4	7	1	0	**19**	11	6	5	0	0	49	0.39
5	2	0	2	7	**12**	2	1	0	0	26	0.46
6	4	1	3	2	2	**7**	0	1	0	20	0.35
7	3	5	0	3	1	0	**10**	0	0	22	0.45
8	0	1	1	0	0	0	0	26	0	28	0.93
9	0	0	0	0	0	0	0	0	84	84	1.00
10	0	0	0	2	2	2	1	3	5	15	
Total	62	21	15	47	43	20	20	30	89	347	
Producer's accuracy	0.63	0.24	0.13	0.40	0.28	0.35	0.50	0.87	0.94		
Overall accuracy for 7 forest types	0.412281										
Kappa accuracy for 7 forest types	0.333343										
Kappa variance for 7 forest types	0.001281										

Table 8.7 Error matrix of sarules4

	Reference Data										
Classified Data	1	2	3	4	5	6	7	8	9	Total	User's accuracy
1	**40**	2	5	13	9	1	1	0	0	71	0.56
2	4	**8**	2	0	1	0	2	0	0	17	0.47
3	3	2	**2**	0	3	2	0	0	0	12	0.17
4	7	2	0	**22**	10	5	7	0	0	53	0.42
5	2	0	2	6	**17**	2	1	0	0	30	0.57
6	3	1	3	3	3	**10**	0	1	0	24	0.42
7	3	5	0	3	0	0	**9**	0	0	20	0.45
8	0	1	1	0	0	0	0	29	0	31	0.94
9	0	0	0	0	0	0	0	0	87	87	1.00
10	0	0	0	0	0	0	0	0	2	2	
Total	62	21	15	47	43	20	20	30	89	347	
Producer's accuracy	0.65	0.38	0.13	0.47	0.40	0.50	0.45	0.97	0.98		
Overall accuracy for 7 forest types	0.473684										
Kappa accuracy for 7 forest types	0.401001										
Kappa variance for 7 forest types	0.001364										

Table 8.8 Error matrix of sarules7

Classified Data	Reference Data 1	2	3	4	5	6	7	8	9	Total	User's accuracy
1	**36**	2	6	14	9	1	1	0	0	69	0.52
2	4	**10**	2	0	1	1	2	0	0	20	0.50
3	5	2	**2**	0	2	1	0	0	0	12	0.17
4	7	1	0	**21**	12	7	5	0	0	53	0.40
5	2	0	2	4	**15**	2	5	0	0	30	0.50
6	3	1	2	3	3	**7**	0	0	0	19	0.37
7	3	5	0	3	0	0	**7**	0	0	18	0.39
8	2	0	1	0	0	0	0	30	0	33	0.91
9	0	0	0	0	0	0	0	0	87	87	1.00
10	0	0	0	2	1	1	0	0	2	6	
Total	62	21	15	47	43	20	20	30	89	347	
Producer's accuracy	0.58	0.48	0.13	0.45	0.35	0.35	0.35	1.00	0.98		
Overall accuracy for 7 forest types	0.429825										
Kappa accuracy for 7 forest types	0.352956										
Kappa variance for 7 forest types	0.001326										

The classification map of fes12 (Plate 31) could not predict the power line easement. Durras Lake and Brush Island were classified as the unknown class, and Willinga River was predicted in a very small part. The classification has also significantly under-estimated the Dry Sclerophyll forest and the Wet *E. maculata* forest and over- estimated the Lower slope wet forest and the Dry *E. maculata* forest. The classification maps of sarules (Plate 32), sarules4 (Plate 33), and sarules7 (Plate 34) are quite similar to one another but different from those of the three individual AI models and the combined AI models. They have predicted Durras Lake nicely. Brush Island was reasonably classified as the unknown class. Willinga River was hardly predicted, as was the power line easement. Comparing with the classification maps of the three individual AI models and the combined AI models, the three classification maps are a bit more fragmented, and they all have under-predicted the Dry *E. maculata* forest. Moreover, there is a proportion of the study area unclassified, among which some islands including Brush Island, the front beaches, and some high land areas are most noticeable.

The Fuzzy Expert Systems could produce maps of certainty factor (e.g., CFc). One such example is given in Plate 35, which displays the certainty factor of the classification map of sarules4. It shows that most of the area has moderate to high certainty factors. It should be noted that this map of certainty factor is not the kind of confidence map illustrated in the next chapter (e.g., Plate 36 and Plate 37) which requires cross-examination of several classifications.

8.2.2 Outputs of the user interface and explanation machine

The Fuzzy Expert Systems implemented the explanation machine through the user interface. Appendix 1 displays the dialogues between a user and the system through the user interface during a typical run. In the dialogues, italic words are FuzzyCLIPS commands entered by the user, bold words are the user's inputs to the system, and others are the responses of the system. It can be seen that when entering data by keyboard, the user could enter either fuzzy or crisp data value (facts). The user could also enter 'why' to get an explanation from the system about the usefulness of a particular input variable (e.g., the sentence that is underlined in Appendix 1). In addition, the user could enter 'NP' when there is no data available for a particular input variable, in which case the system might still give a correct answer. Meanwhile, the '*(agenda)*' command would list all rules that partially match the input facts. Firing the rules would display the rule names, the conclusions and their associated CF_cs. The conclusion with the highest CF_c would be considered to be the final answer, and the rule that inferred the conclusion could be displayed to explain the reasoning process behind the conclusion. To answer the "what if…" question, the system would simply ask the user to continue and to enter new data values, which would deduce a new conclusion. After a conclusion is inferred, the system would require the user to judge the answer. If the answer was considered to be correct, the user could choose to continue or to quit. However, if the answer was judged to be incorrect, the system would ask for the right answer from the user, and then it would create a new rule in the rule base. Therefore, at the next time, when the same data values had been input, the system would give the correct answer by firing the newly created rule. Generally speaking, the user interface is user friendly. It has provided sufficient help information, and it is easy to follow.

8.3 DISCUSSION

The study has shown that the Fuzzy (Rule-based) Expert Systems can be used for complicated forest type predictive mapping with fair predictive accuracies. One major advantage of the Fuzzy Expert Systems is their comprehensibility, which represents the classification processes as production rules of natural language instead of decision trees or network connections. On the other hand, one major disadvantage of the Fuzzy Expert Systems is the time and resource requirement in building the fuzzy rule bases and implementing the classifications to the whole study area. Therefore, using the Fuzzy Expert Systems for forest type mapping is a trade-off between efficiency and comprehensibility.

Generating the fuzzy rule bases were a vital part of these Fuzzy Expert Systems. This study has largely escaped the knowledge acquisition "bottleneck" problem by learning the classification rules directly from samples, which was recommended by the study of Wang and Mendel (1992). Thus, the author did not spend months and years in looking for domain experts, preparing questionnaires, interviewing domain experts, and resolving disagreement between domain experts. In fact, this study took only around one month man power to generate the fuzzy rule bases. However, one assumption for the effectiveness of the method is that the learning samples obtained from the field survey or the existing classification maps are reliable and widely representative.

Another feature of the Fuzzy Expert Systems is that they used fuzzy logic instead of the traditional Boolean logic for their inference engines. The advantage of the fuzzy reasoning relies on its effectiveness in handling the classification uncertainty. For example, in this study, it is more reasonable to assume that any pixel may at the same time belong to several classes with different degrees, in which the fuzzy reasoning is more appropriate. On the other hand, Boolean logic could not deal with the issue of partial matching, which would easily cause large unclassified areas in the classification maps. However, the potentially large number of partial matchings has significantly increased the time and resource requirement (See Table 8.4). This is an important limiting factor for the application of the Fuzzy Expert Systems when compared with the above three individual AI models and the combined AI models.

The user interface of the Fuzzy Expert Systems is relatively simple. Though it has not provided many comprehensive functions, it is adequate for most users. It may be disappointing that window based user interface could not be created due to the inherent limitation of FuzzyCLIPS language. However, there is room for further improvement. For example, improvement can be made in coping with unexpected user input and in explaining reasoning processes.

The Fuzzy Expert System classification of fes12 was totally unsatisfying in terms of predictive accuracies, visual appearance, and the time and resource requirement. One reason for the disappointing performance of fes12 is that the absolute size of its learning samples is too small. Another reason is that the base map of D-S1 is not that reliable with an overall predictive accuracy of only over 50% for the seven forest types. In addition, selecting only the most representative samples from each class has ignored the large within class variance. On the other hand, the size of the learning samples selected from the field samples (i.e., sarules) was more than doubled. It has been demonstrated that this did fix the problems and largely improved the classification performance. The slightly better Kappa accuracy of sarules over that of the Artificial Neural Network is encouraging. Meanwhile, by resolving the unknown class from sarules sarules4 has largely increased the predictive accuracies. More importantly, it did reduce the total number of unclassified pixels (see Figure 8.7). Moreover, pruning the rule base of sarules4 to sarules7 has proved to be worthwhile. Even though this has decreased the predictive accuracies to some extent and has increased the number of unclassified pixels (Figure 8.7), it did reduce the size of the rule base significantly, which has provided better generalizations and comprehensibility for the classification.

The three quite similar classification maps of sarules, sarules4, and sarules7 have indicated that the differences among the three Fuzzy Expert Systems are not significant. It is not clear why the three Fuzzy Expert Systems produced more fragmented classification maps. The 5%-7% unclassified area on these maps is not significant, and this does not always represent failure. Sometimes, this could actually represent having captured desirable classifications, especially in cases when the pixels could not be comfortably classified into any of the nine classes. For example, classifying the several islands into the unknown class is quite reasonable. In addition, classifying the sand beaches into the unknown class is also understandable, as the beaches have quite different characteristics from the Clear land/Paddock class. The map of certainty factor provided another aspect of the classification, which has similar meaning as the probability map of other classifiers.

8.4 SUMMARY

In summary, the experiment of Fuzzy Expert System was successful. The three Fuzzy Expert Systems were capable classifiers for complicated forest type mapping. They have produced comparable classifications with the three individual AI models. The trade off is between the efficiency and the comprehensibility. Building the fuzzy rule bases directly from the learning samples is the most important attractive feature of these Fuzzy Expert Systems. Other attractive features include the fuzzy inference engines and the simple user interface and explanation machine.

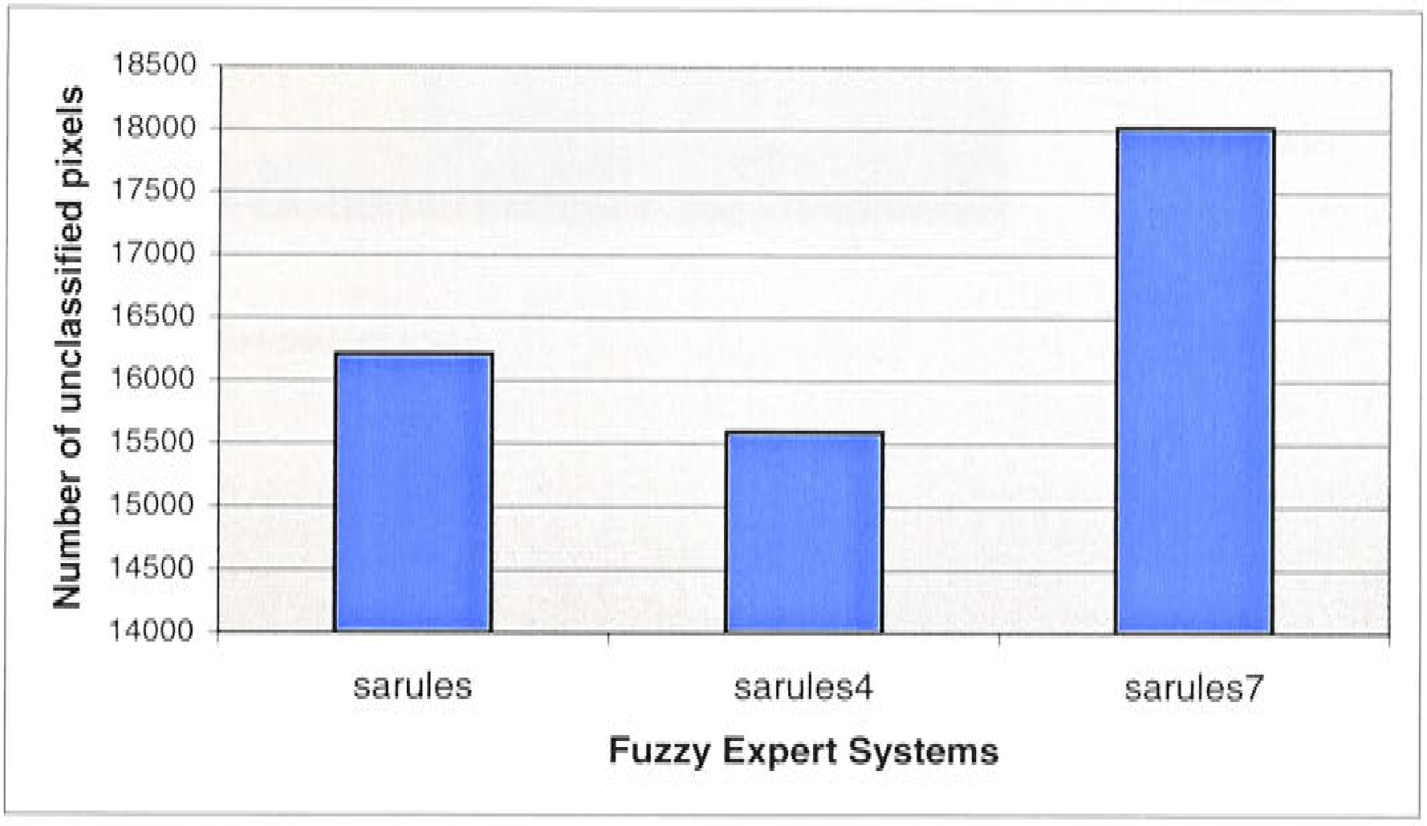


Figure 8.7 Number of unclassified pixels in the maps of sarules, sarules4, and sarules7

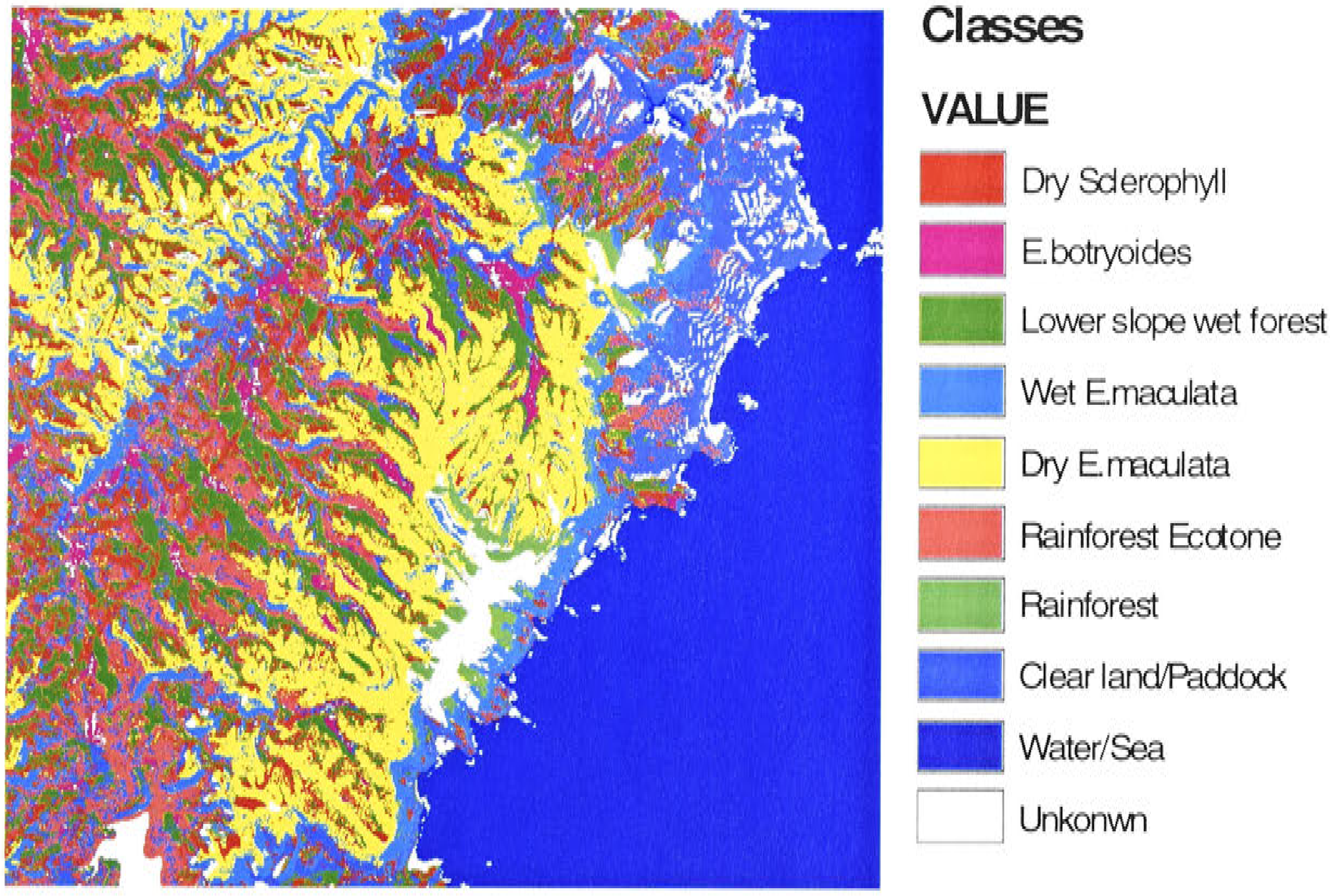


Plate 31 Classification map of fes12

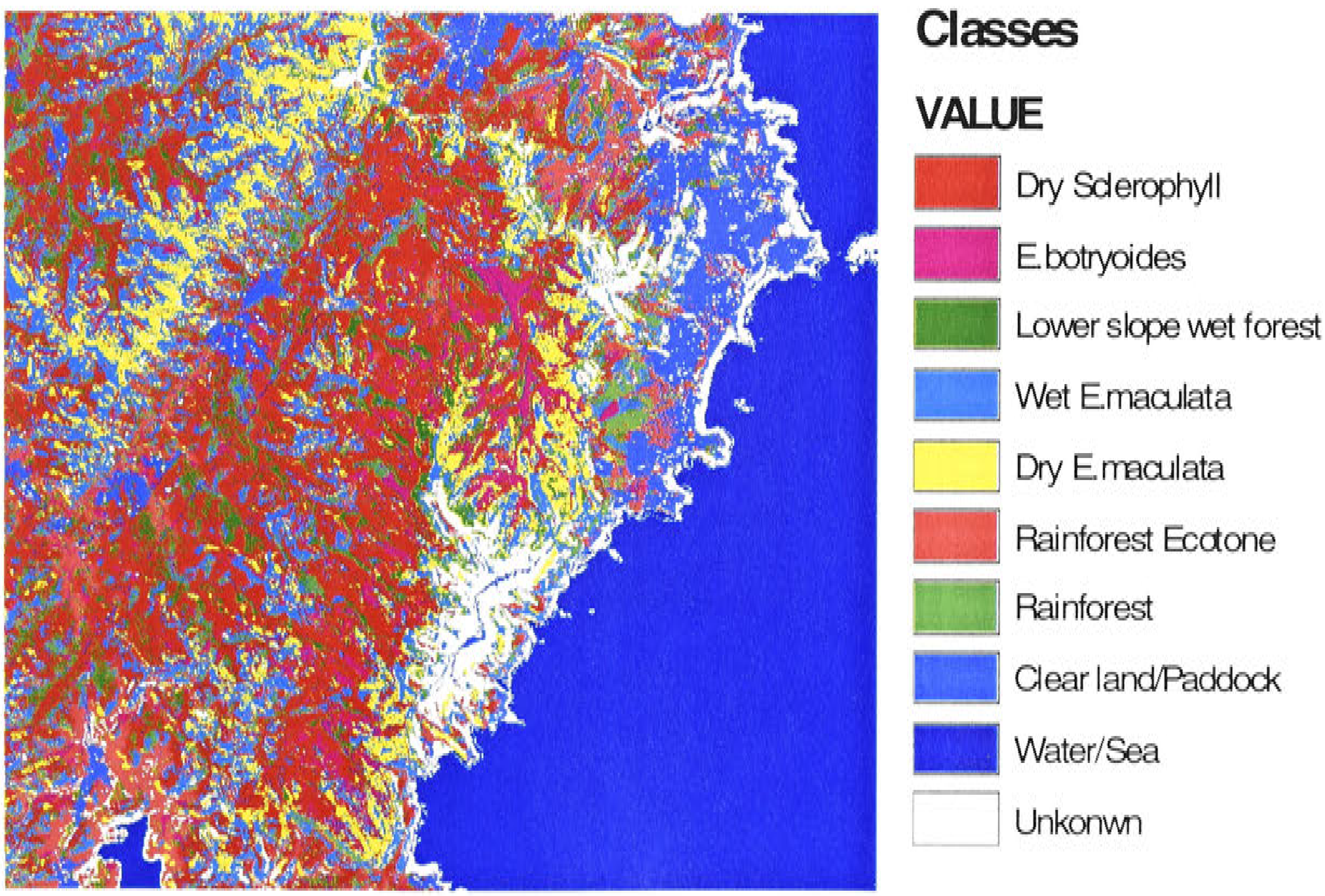


Plate 32 Classification map of sarules

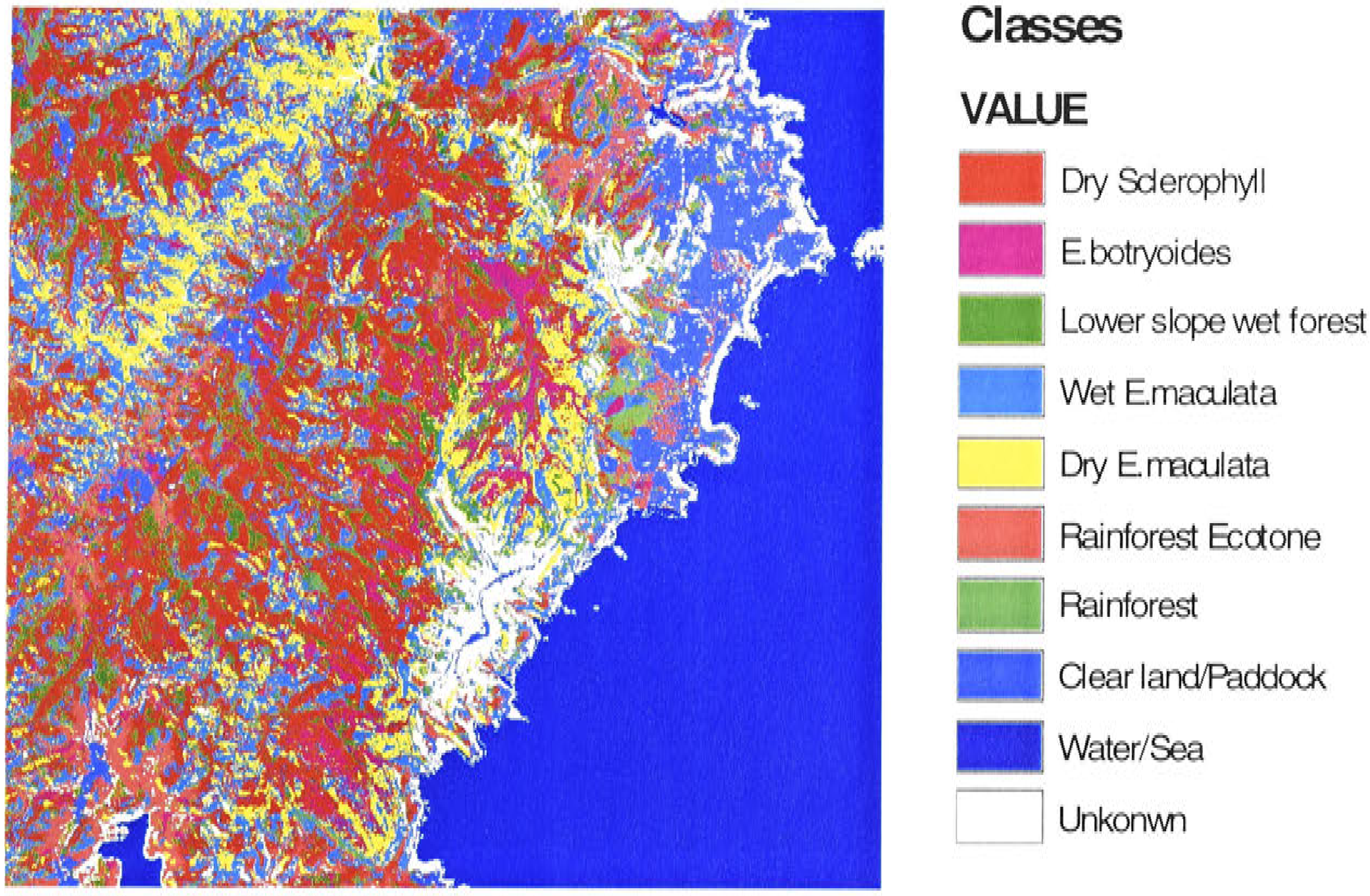


Plate 33 Classification map of sarules4

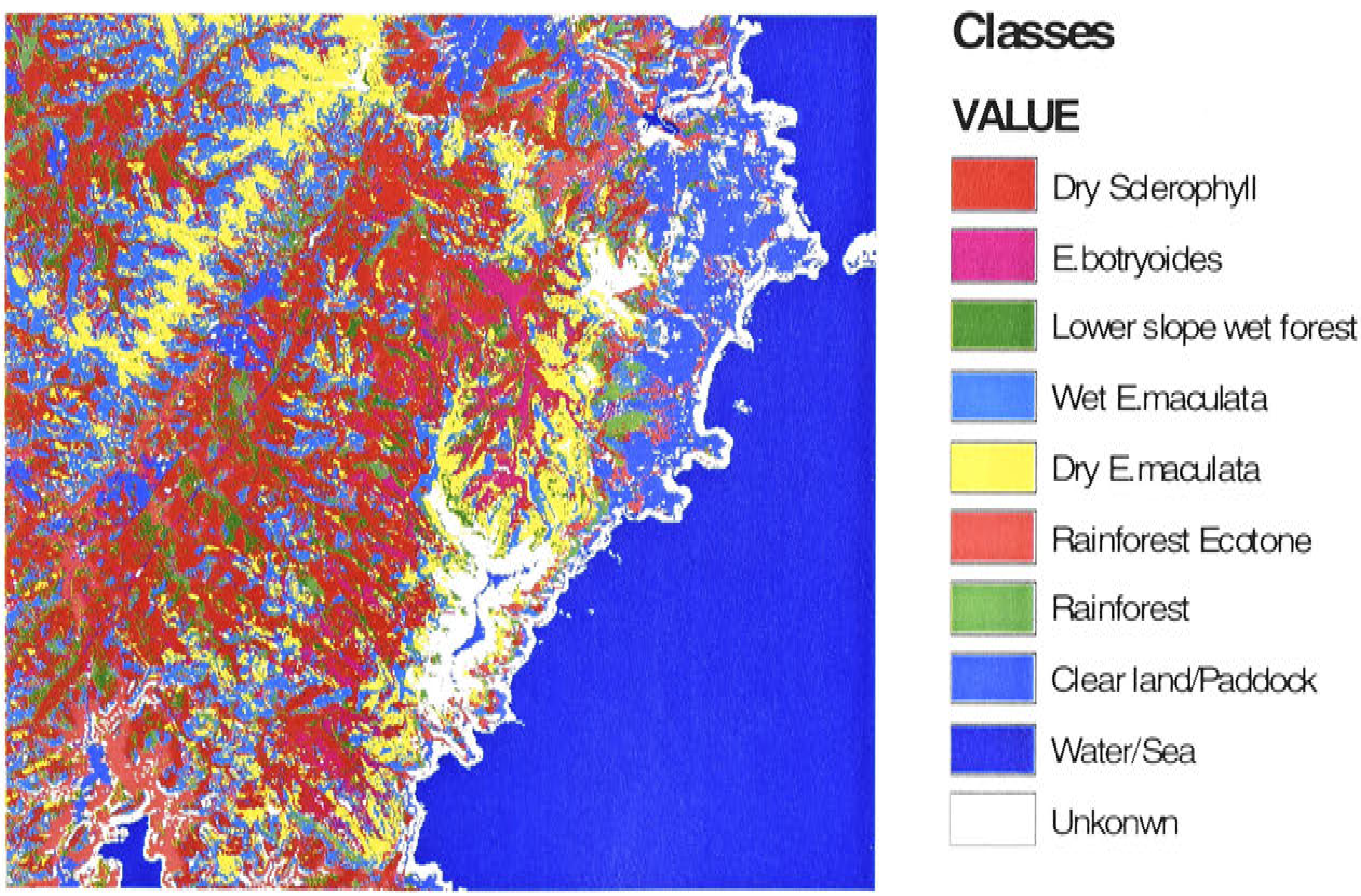


Plate 34 Classification map of sarules7

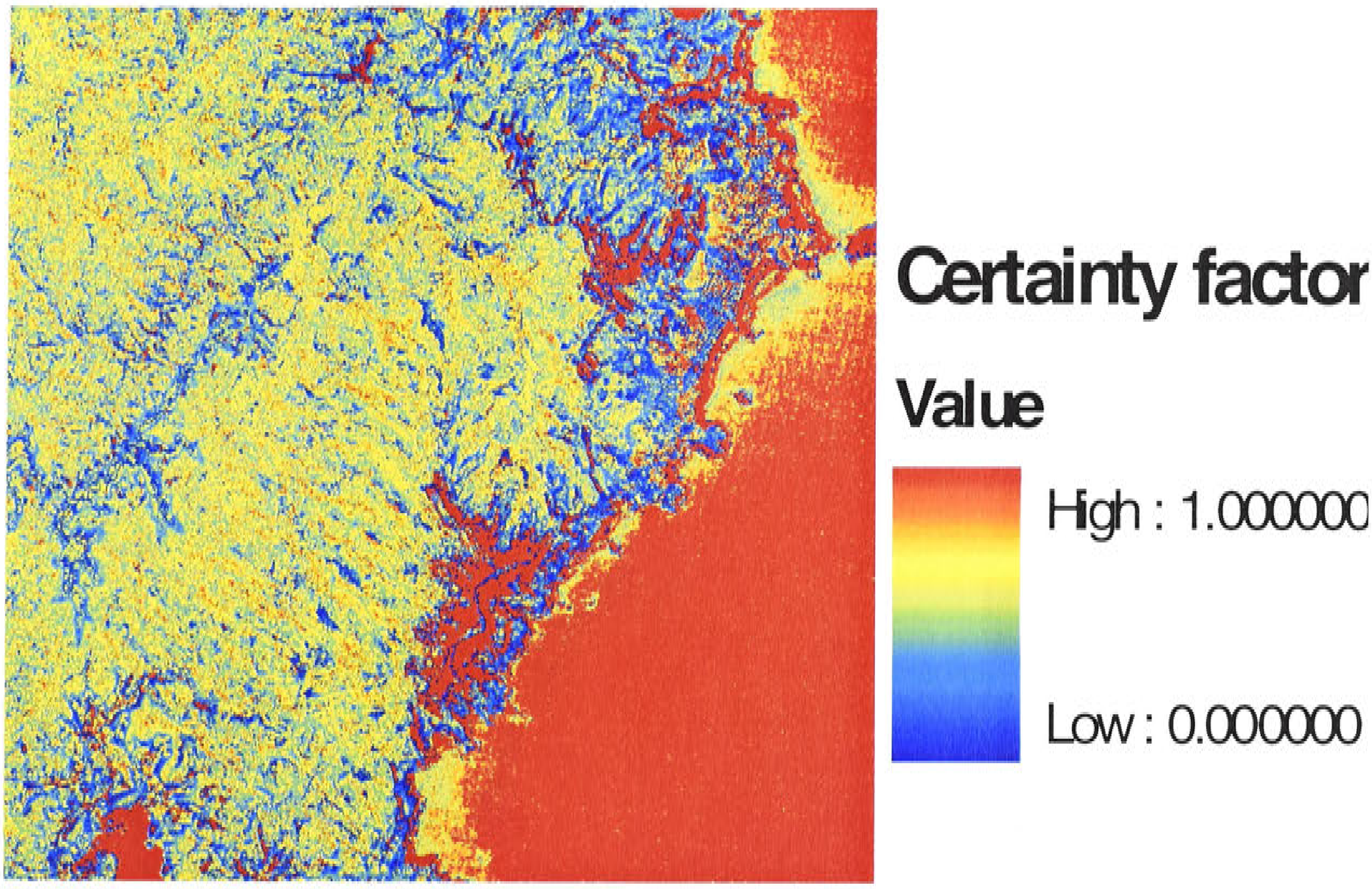


Plate 35 Certainty factor map of sarules4

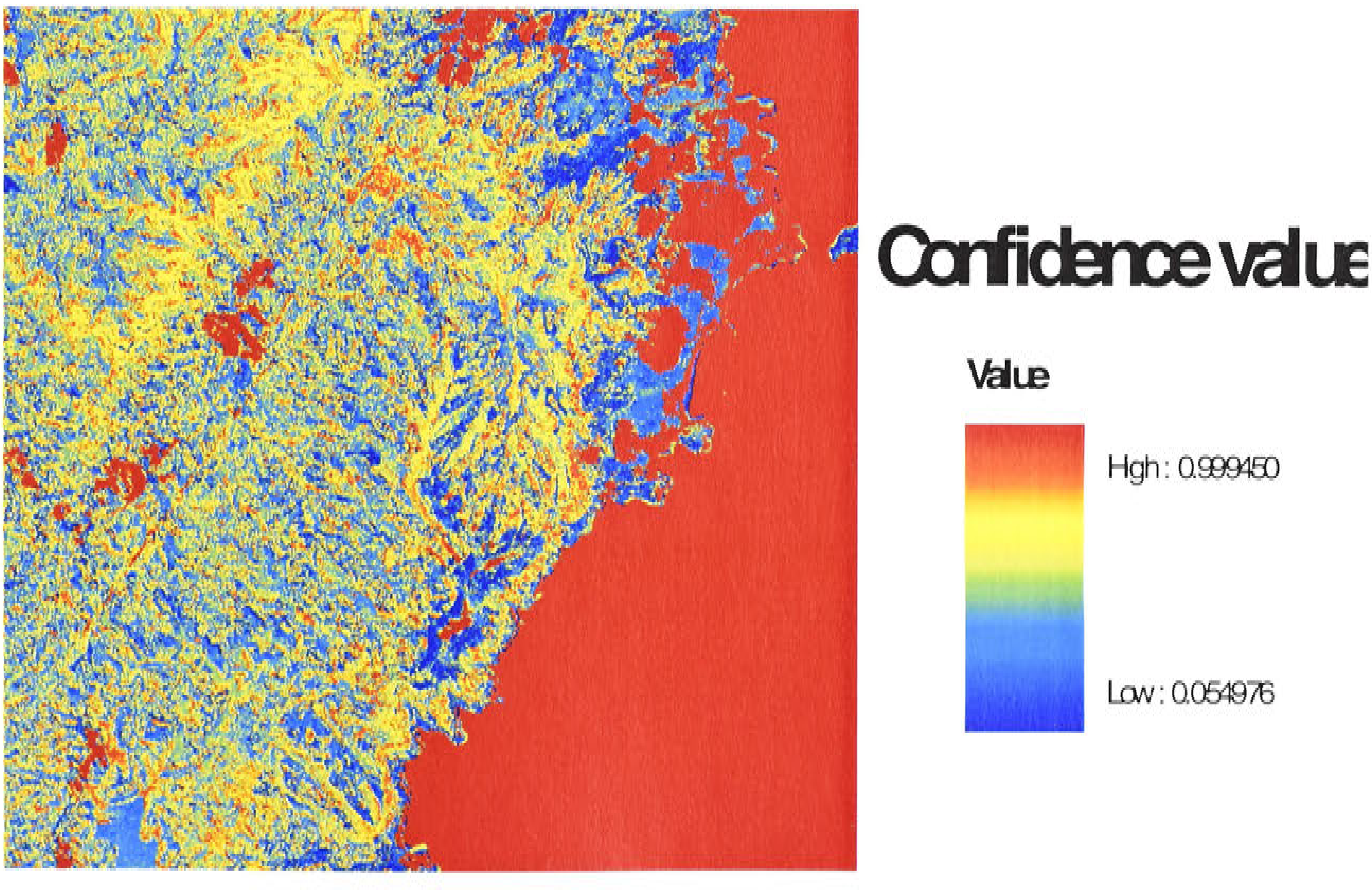


Plate 36 Quantitative confidence map of D-S1

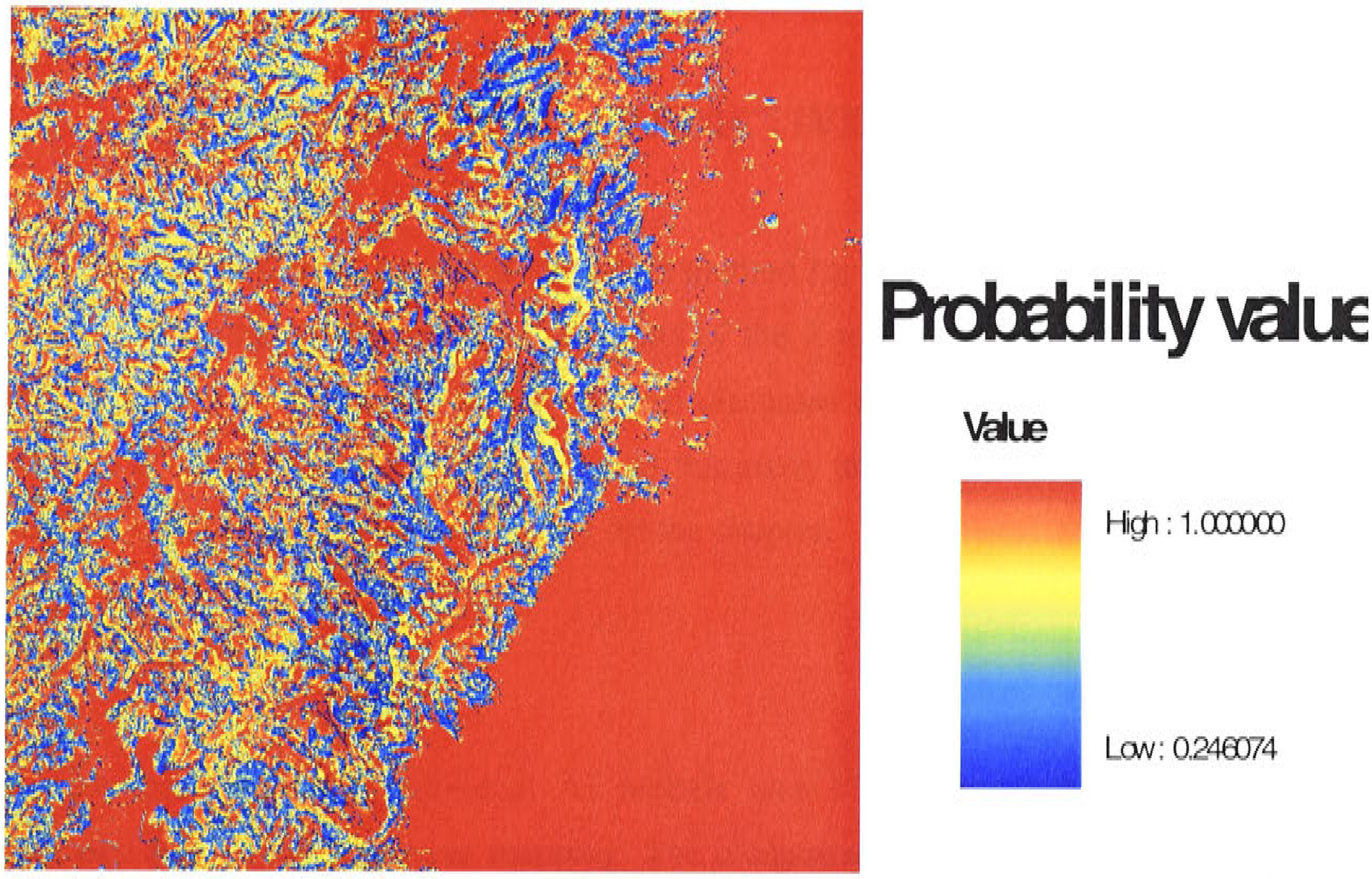


Plate 37 Probability map of D-S1

Chapter 9

COMPARISONS OF PREDICTIVE MODELS

The chapter first describes the statistical Z-test method for comparing the classifiers of this study. Then, the chapter reports the results of the Z-test for the 34 individual and combined AI models applied in this study, and discusses the findings. Following that, the chapter compares the predictive models of this study and those of previous studies. Finally, the chapter gives a short summary for the chapter.

9.1 METHOD OF Z-TEST

This study has applied 7 individual AI models and 27 combined AI models for complicated forest type predictive mapping. The comparisons of these classifiers have been made through the traditional accuracy assessment and visual assessment. A general impression is that the combination strategy did make significant improvements over the initial classifications. To test the impression statistically, the Z-test (Cohen, 1960) was used to compare each classifier against others to see whether or not they are significantly different.

Cohen (1960) introduced the Z-test to test the significance of the difference between two independent classifiers. The formula is:

$$Z = \frac{K_1 - K_2}{\sqrt{\sigma_{K_1}^{\ 2} + \sigma_{K_2}^{\ 2}}} \tag{9.1}$$

where K_1 and K_2 are the Kappa accuracies of two classifiers, and σ_{K_1} and σ_{K_2} are the variances associated with the K_1 and K_2. Cohen (1960) originally suggested a formula to calculate the approximate large sample variance of Kappa later shown to be incorrect (Rosenfield & Fitzpatrick-Lins, 1986). The correct one is given by Bishop *et al.* (1975, p 396), which is as follows:

$$\sigma_K^{\ 2} = \frac{1}{N} \times \left\{ \frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^{\ 2})}{(1-\theta_2)^4} \right\} \tag{9.2}$$

$$\text{where } \theta_1 = \sum_{i=1}^{k} P_{ii} = \frac{\sum_{i=1}^{k} X_{ii}}{N}, \tag{9.3}$$

$$\theta_2 = \sum_{i=1}^{k} P_{i+}P_{+i} = \frac{\sum_{i=1}^{k} X_{i+}X_{+i}}{N^2}, \tag{9.4}$$

$$\theta_3 = \sum_{i=1}^{k} P_{ii}(P_{i+} + P_{+i}) = \frac{\sum_{i=1}^{k} X_{ii}(X_{i+} + X_{+i})}{N^2}, \tag{9.5}$$

$$\theta_4 = \sum_{i=1}^{k}\sum_{j=1}^{k} P_{ij}(P_{j+} + P_{+i})^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} X_{ij}(X_{j+} + X_{+i})^2}{N^3} \tag{9.6}$$

However, to calculate the $\sigma_K{}^2$ on the 7 forest types from the error matrix of total 9 classes, the above formulas must be modified as follows:

$$\sigma_K{}^2 = \frac{1}{N_1} \times \left\{ \frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2{}^2)}{(1-\theta_2)^4} \right\} \tag{9.7}$$

$$\text{where } \theta_1 = \sum_{i=1}^{7} P_{ii} = \frac{\sum_{i=1}^{7} X_{ii}}{N_1}, \tag{9.8}$$

$$\theta_2 = \sum_{i=1}^{7} P_{i+}P_{+i} = \frac{\sum_{i=1}^{7} X_{i+}X_{+i}}{\frac{N}{N_1}}, \tag{9.9}$$

$$\theta_3 = \sum_{i=1}^{7} P_{ii}(P_{i+} + P_{+i}) = \frac{\sum_{i=1}^{7} X_{ii}(X_{i+} + X_{+i})}{\frac{N}{N_1}}, \tag{9.10}$$

$$\theta_4 = \sum_{i=1}^{7}\sum_{j=1}^{7} P_{ij}(P_{j+} + P_{+i})^2 = \frac{\sum_{i=1}^{7}\sum_{j=1}^{7} X_{ij}(X_{j+} + X_{+i})^2}{\frac{N^2}{N_1}} \tag{9.11}$$

where $N_1 = \sum_{i=1}^{7} X_{+i}$ is the total number of field samples for the 7 forest types (228 in this case).

9.2 RESULTS OF Z-TEST AND DISCUSSION

The Z-test, which is obtained from the error matrices, can provide useful information in comparing the classifiers. The results of the Z-test for the total 34 classifiers are listed in Table 9.1. Two confidence levels at 90% (i.e., Z statistic > 1.645) and 95% (i.e., Z statistic > 1.96) were evaluated from the table.

The following findings can be implied from the Z-test:

- Differences among the Decision Tree, the Artificial Neural Network, and the model based on Dempster-Shafer's theory are not significant.
- Differences among the Fuzzy Expert Systems of sarules, sarules4, and sarules7 are not significant.
- Differences among the above six individual AI models are not significant.
- Differences among all of the combined AI models are not significant except that there is significant difference between combine1 and vote7 at the 90% confidence level.
- The Fuzzy Expert System of fes12 is significantly different from all other AI models at the 95% confidence level.
- The Fuzzy Expert System of sarules is significantly different from six of the combined AI models at the 95% confidence level and the other four of the combined AI models at the 90% confidence level.
- The Fuzzy Expert System of sarules4 is not significantly different from any combined AI model.
- The Fuzzy Expert System of sarules7 is significantly different from two of the combined AI models at the 95% confidence level and the other two of the combined AI models at the 90% confidence level.
- The individual model based on Dempster-Shafer's theory is significantly different from ten of the combined AI models at the 95% confidence level and the other 12 of the combined AI models at the 90% confidence level.
- The individual model of Artificial Neural Network is significantly different from 9 of the combined AI models at the 95% confidence level and the other 10 of the combined AI models at the 90% confidence level.
- The combined AI model of vote7 at second stage is significantly different from the individual model of Decision Tree at the 90% confidence level.

The findings have shown that most of the combined AI models are significantly different from the Artificial Neural Network and the model based on Dempster-Shafer's theory. More impressive, the difference between vote7, which is the best combined AI model from the two stages' combinations, and the Decision Tree, which is the best individual AI model entering the combination, is significant at the 90% confidence level. This is an important achievement that this study has managed to obtain, and it confirms that the general hypothesis of this study is correct. The test has confirmed that the classification of fes12 is completely unsuccessful. It has also shown that other 6 individual AI models were comparable classifiers.

9.3 COMPARISONS OF THE PREDICTIVE MODELS OF THIS STUDY WITH THOSE OF PREVIOUS STUDIES

The study area, as reviewed in Chapter 3, has attracted a large number of studies. Many of these studies used AI models for land cover classification. The results of these previous studies are not strictly comparable with those of this study. Because even though they may use the same data set, other factors such as the different partition of training and test sets will have impacts on classification accuracy. Moreover, the resultant classification accuracy is not the only assessment criterion for classifiers. Criteria such as computation efficiency, ease of use of a model, visual appearance, and comprehensiveness of the modelling results should also be considered in assessing a classifier. However, a reasonable comparison of different studies based on only classification accuracy is informative and valuable. The following paragraphs give such a comparison of predictive models of this study with those of previous studies based on their overall classification accuracies. We then draw some conclusions from the findings.

Several Decision Trees and Artificial Neural Networks have been applied for predictive forest mapping in the study area (see section 3.3). The classification accuracies of these models are varied. The reported overall classification accuracies of the predictive models of this study appear to be relatively low. This is because that this study has focused on assessing the predictive accuracies only for the 7 forest types due to the identified sampling error associated with the other two classes (e.g., the Water/Sea class and the Clear land/Paddock class). To facilitate comparison, the classification accuracies of 18 predictive models of this study and previous studies are summarised in

Table 9.2. It should be noted that these accuracies are either reported on the papers or calculated from the error marties that shown on the papers. Those with a superscript "*" are training accuracies, since the test accuracies are not reported.

Table 9.2 Classification accuracies of 18 models of this study and previous studies

Study	Predictive Model	Overall Accuracy for the 7 forest types	Overall Accuracy for all 9 classes
Fitzgerald and Lees, 1992	Backpropagation ANN	48.8%*	NA
Fitzgerald and Lees, 1994	Backpropagation ANN with 3 * 3 windows	NA	57.1%*
Fitzgerald and Lees, 1994	Backpropagation ANN with 5 * 5 windows	35.9%*	50.6%*
Lees and Ritman, 1991	CART Decision Tree	49.1%*	NA
Gahegan and West, 1998	DONNET ANN	NA	70-75%
German, 1999	DONNET ANN	NA	72.6%
German, 1999	C4.5 Decision Tree	NA	67%
German, 1999	Vanilla Backpropagation ANN	NA	50.8%
German, 1999	Restricted DONNET ANN	NA	75.4%
Gahegan and Takatsuka, 1999	SOM ANN + LVQ ANN	NA	68.6%
This study	(CART) Decision Tree	65%* / 46.5%	78.3%* / 63.7%
This study	(Backpropagation) Artificial Neural Network	53.1%* / 41.2%	71.7%* / 60.8%
This study	Dempster-Shafer's theory	44.5%* / 39%	66.3%* / 58.5%
This study	D-S1	51.3%	66.9%
This study	ffamaxclass	51.3%	66.9%
This study	fameanppclass	49.6%	66.3%
This study	Vote7	53.9%	68.3%
This study	Sarules7	43%	62%

* These are training accuracies, others are test accuracies. NA: the accuracy is not available.

The following findings can be implied from Table 9.2:

- The earlier studies of Fitzgerald, Lees, and Ritman used simple backpropagation ANNs or CART Decision Tree which resulted in relatively low overall accuracies.
- The more recent studies of Gahegan, German, West, and Takatsuka used advanced ANNs or C4.5 Decision Tree which resulted in largely improved overall accuracies.
- This study used simple individual AI models and combined models which resulted in overall accuracies relatively lower than the studies of Gahegan, German, West, and Takasuka but higher than the studies of Fitzgerald, Lees, and Ritman.
- The backpropagation ANN used in this study obtained higher overall accuracy than those used in Fitzgerald and Lees' studies (e.g., 53.1% Vs 48.8% and 35.9%; 71.7% Vs 57.1% and 50.6%).
- The CART Decision Tree used in this study obtained much higher overall accuracy than that used in Lees and Ritman's study (e.g., 65% Vs 49.1%).

- The CART Decision Tree used in this study resulted in lower overall accuracy than the C4.5 Decision Tree used in German's study (e.g., 63.7% Vs 67%).
- The backpropagation ANN used in this study resulted in higher overall accuracy than the vanilla backpropagation ANN used in German's study (e.g., 60.8% Vs 50.8%).
- The backpropagation ANN used in this study obtained much lower overall accuracy than the DONNET ANNs used in Gahegan, German, and West studies (e.g., 60.8% Vs 70-75%).
- The backpropagation ANN used in this study resulted in lower overall accuracy than the combined SOM and LVQ ANNs used in Gahegan and Takatsuka's study (e.g., 60.8% Vs 68.6%).
- The best combined AI model of this study – vote7 resulted in comparable overall accuracy with the combined SOM and LVQ ANNs used in Gahegan and Takatsuka's study (e.g., 68.3% Vs 68.6%).
- The best combined AI model of this study – vote7 resulted in lower overall accuracy than the DONNET ANNS used in Gahegan, German, and West studies (e.g., 68.3% Vs 70-75%).
- The best combined AI model of this study – vote7 resulted in higher overall accuracy than the C4.5 Decision Tree used in German's study (e.g., 68.3% Vs 67%).

It is interesting to find that the CART Decision Tree used in this study obtained lower overall accuracy than the C4.5 Decision Tree used in German's study. The conclusion, however, can not be drawn that C4.5 Decision Tree is better than CART Decision Tree. Because the CART Decision Tree did not apply a pruning process while the C4.5 Decision Tree did. More importantly, the two Decision Trees have different splitting rules as reviewed in Chapter 2, and therefore may be suitable for different applications. More studies are needed to compare the performance of the two Decision Trees under a same environment before a more conclusive recommendation can be reached.

On the other hand, the much better classification performance of DONNET and restricted DONNET than that of backpropagation ANN is understandable. Because DONNET is a more complicated and advanced ANN, and it starts not from random network weights but from a discriminate analysis. In addition, this study has not yet explored the full potential of backpropagation ANN, as no parameter and structure optimisation or fine-tuning process has been applied in this study.

It is a disappointment that the best combined AI model of this study – vote7 did not achieve the same level classification accuracy as the DONNET ANN. The reason is that it is build on the three simple individual AI models that achieved 10% or so lower overall accuracies than the DONNET ANN. Therefore not a jump of accuracy improvement should be expected by any combination process. This finding doesn't weaken our conclusions on combining models. On the contrary, the author believes that combining individual models of different principles would always improve classification performance. Thus, if for example a DONNET ANN, a C4.5 Decision Tree, and a third model such as genetic algorithm are to be combined using the methods applied and developed in this study, the author believes that it would increase classification accuracy to some extent. Meanwhile, this study has focused on the prediction of the 7 forest types. Unable to compare the predictive accuracy for the 7 forest types between vote7 and DONNET ANN is disappointing. The author believes that the predictive accuracy of vote7 for the 7 forest types is still lower than those of DONNET ANNs. However, the author suspects that the extent of difference may not be as large as that between the predictive accuracies of the 9 classes. One fact is that vote7 increased 12.7% predictive accuracy for the 7 forest types but only 7.5% predictive accuracy for the 9 classes when compared to the backpropagation ANN.

Finally, creating a super model for predictive forest mapping was never an intention of this study. The objective of this study is to demonstrate that combining models is a valuable practice that has several advantages including increase of prediction accuracy. When assessing a classifier, not only its predictive accuracy but also other criteria such as computation requirement and comprehensiveness should also be taken into account. In selecting a classifier, it often involves balancing trade-offs among these criteria. For example, sarules7 of this study has advantage of comprehensiveness, even though its prediction is much poorer than the DONNET ANN.

9.4 SUMMARY

The Z-test indicates that the combination strategy did make a significant difference from the individual AI models for forest type mapping. This reinforces our confidence that the combination strategy can improve classification performance in many circumstances.

Meanwhile, the comparison of the predictive models of this study and those of previous studies in terms of overall classification accuracy shows that the combination models of this study are better predictive models than those simple AI models used in earlier studies of Fitzgerald, Lees, and Ritman. But they are worse than those advanced AI models used in more recent studies of Gahegan, German, West, and Takatsuka. Nevertheless, the findings do not weaken our conclusions on combining models. Because combining models have several advantages over individual models besides increasing classification accuracy.

Table 9.1 Z-test of all AI models applied in this study

ZTEST	vote7	ffamaxclass	fameanppclass	d-s1	combine2	combine1	meanclass	d-s	ann	dt	medianclass	maxclass	fmeanclass	fmeanppclass	fameanclass	fmedianclass	fmedppclass	famedianclass	famedppclass	fmaxclass	fmaxppclass	famaxclass	famaxpclass	fcmaxclass	fffmaxclass	fffamaxclass	ffagreeclass	ffappclass	vote1	vote6	fes12	sarules4	sarules	sarules7
vote7	-	0.546	0.935	0.582	1.011	**1.748**	0.863	***2.924***	***2.866***	**1.693**	1.140	1.209	1.232	1.242	1.320	1.413	1.310	1.128	1.151	1.106	1.196	1.196	1.010	0.768	0.835	0.645	1.323	1.050	0.270	0.193	***6.027***	1.295	***2.621***	***2.224***
ffamaxclass		-	0.390	0.036	0.465	1.200	0.318	***2.366***	***2.311***	1.149	0.594	0.664	0.687	0.697	0.775	0.868	0.764	0.582	0.607	0.561	0.651	0.651	0.465	0.223	0.290	0.100	0.778	0.505	0.275	0.352	***5.434***	0.747	***2.066***	**1.673**
fameanppclass			-	0.354	0.075	0.807	0.072	***1.966***	**1.914**	0.760	0.203	0.274	0.298	0.308	0.385	0.477	0.374	0.191	0.217	0.171	0.262	0.262	0.075	0.166	0.100	0.290	0.388	0.116	0.665	0.742	***5.005***	0.355	**1.668**	1.278
d-s1				-	0.429	1.164	0.282	***2.330***	***2.275***	1.113	0.558	0.628	0.651	0.661	0.739	0.832	0.728	0.546	0.570	0.525	0.615	0.615	0.429	0.187	0.254	0.064	0.742	0.469	0.312	0.389	***5.395***	0.711	***2.030***	1.637
combine2					-	0.732	0.147	**1.890**	**1.838**	0.685	0.128	0.199	0.223	0.233	0.310	0.402	0.299	0.116	0.142	0.096	0.187	0.187	0.000	0.242	0.175	0.365	0.313	0.041	0.740	0.817	***4.925***	0.279	1.592	1.203
combine1						-	0.880	1.153	1.104	0.040	0.604	0.532	0.507	0.497	0.419	0.328	0.432	0.615	0.588	0.634	0.543	0.543	0.732	0.974	0.907	1.098	0.416	0.689	1.476	1.553	***4.160***	0.454	0.857	0.471
meanclass							-	***2.041***	***1.988***	0.832	0.275	0.346	0.370	0.380	0.457	0.550	0.446	0.264	0.289	0.243	0.334	0.334	0.147	0.094	0.028	0.218	0.460	0.188	0.593	0.670	***5.086***	0.427	**1.742**	1.352
d-s								-	0.043	1.183	**1.762**	**1.687**	**1.659**	**1.650**	1.572	1.480	1.585	**1.773**	**1.743**	**1.790**	**1.697**	**1.697**	**1.891**	***2.135***	***2.067***	***2.261***	1.569	**1.845**	***2.647***	***2.724***	***2.987***	1.614	0.297	0.680
ann									-	1.134	**1.710**	1.635	1.608	1.599	1.521	1.429	1.535	**1.721**	**1.691**	**1.738**	**1.646**	**1.646**	**1.838**	***2.081***	***2.014***	***2.206***	1.518	**1.793**	***2.590***	***2.668***	***3.014***	1.563	0.252	0.634
dt										-	0.559	0.487	0.462	0.452	0.376	0.285	0.388	0.570	0.543	0.589	0.498	0.498	0.686	0.925	0.859	1.048	0.373	0.643	1.423	1.499	***4.158***	0.410	0.889	0.506
medianclass											-	0.071	0.095	0.105	0.183	0.275	0.171	0.012	0.014	0.032	0.059	0.059	0.128	0.370	0.303	0.493	0.186	0.087	0.869	0.946	***4.792***	0.151	1.464	1.075
maxclass												-	0.024	0.034	0.112	0.203	0.100	0.082	0.057	0.102	0.012	0.012	0.199	0.440	0.373	0.563	0.115	0.158	0.939	1.015	***4.707***	0.080	1.390	1.002
fmeanclass													-	0.010	0.087	0.179	0.076	0.107	0.081	0.127	0.036	0.036	0.223	0.464	0.397	0.587	0.090	0.182	0.962	1.039	***4.673***	0.055	1.363	0.976
fmeanppclass														-	0.077	0.169	0.066	0.117	0.091	0.136	0.046	0.046	0.233	0.474	0.407	0.597	0.080	0.192	0.972	1.049	***4.663***	0.045	1.353	0.966
fameanclass															-	0.092	0.012	0.194	0.168	0.214	0.123	0.123	0.311	0.551	0.485	0.674	0.003	0.269	1.050	1.127	***4.583***	0.032	1.275	0.889
fmedianclass																-	0.103	0.286	0.260	0.306	0.215	0.215	0.403	0.643	0.577	0.767	0.089	0.361	1.143	1.219	***4.488***	0.125	1.183	0.797
fmedppclass																	-	0.183	0.157	0.202	0.112	0.112	0.299	0.540	0.474	0.663	0.015	0.258	1.040	1.116	***4.601***	0.021	1.289	0.901
famedianclass																		-	0.026	0.020	0.071	0.071	0.117	0.358	0.291	0.481	0.197	0.075	0.857	0.934	***4.802***	0.163	1.475	1.086
famedppclass																			-	0.046	0.045	0.045	0.142	0.383	0.316	0.506	0.171	0.101	0.881	0.958	***4.763***	0.137	1.446	1.058
fmaxclass																				-	0.090	0.090	0.097	0.337	0.271	0.461	0.217	0.055	0.836	0.913	***4.816***	0.183	1.493	1.105
fmaxppclass																					-	0.000	0.187	0.428	0.361	0.551	0.126	0.146	0.926	1.003	***4.714***	0.092	1.400	1.013
famaxclass																						-	0.187	0.428	0.361	0.551	0.126	0.146	0.926	1.003	***4.714***	0.092	1.400	1.013
famaxpclass																							-	0.241	0.175	0.365	0.314	0.041	0.740	0.817	***4.926***	0.280	1.593	1.203
fcmaxclass																								-	0.067	0.123	0.554	0.282	0.498	0.575	***5.182***	0.522	**1.836**	1.446
fffmaxclass																									-	0.190	0.488	0.215	0.565	0.642	***5.111***	0.455	**1.769**	1.379
fffamaxclass																										-	0.677	0.405	0.375	0.452	***5.315***	0.646	***1.962***	1.570
ffagreeclass																											-	0.272	1.053	1.130	***4.579***	0.035	1.272	0.886
ffappclass																												-	0.780	0.856	***4.870***	0.238	1.548	1.160
vote1																													-	0.077	***5.731***	1.023	***2.346***	***1.951***
vote6																														-	***5.812***	1.100	***2.423***	***2.028***
fes12																															-	4.648	***3.290***	***3.671***
sarules4																																-	1.316	0.926
sarules																																	-	0.384
sarules7																																		-

Note: Bold numbers are significant at the confidence level of 90%, and bold and italics numbers are significant at the confidence level of 95%.

Chapter 10

CONCLUSIONS

This study aimed to develop effective methods for the mapping of complicated forest types in Kioloa area, NSW. To fulfill the goal, the study first evaluated the effectiveness of some popular AI models in mapping the complicated forest types. The study then attempted to develop a new strategy to combine these AI models and to examine the advantages of the combined AI models in the forest type mapping. The study then wished to find out the modes of the individual AI models and the effectiveness of the combined AI models in handling identified data errors. Finally, the study wanted to develop Fuzzy Expert Systems for forest type mapping and examine their usefulness by comparing them with other individual AI models.

The study was divided into four stages. In the first stage, a Decision Tree, an Artificial Neural Network and a model based on Dempster-Shafer's theory were separately applied to the complicated predictive forest type mapping, and their results were evaluated using traditional accuracy assessment and visual assessment. In the second stage, the outcomes of the three individual AI models were combined using different combination approaches for the complicated forest type mapping. The third stage used methods to evaluate how well the individual and combined AI models handled the known data errors and to examine why the three individual AI models dealt with the data errors so differently. In the fourth stage, four Fuzzy Expert Systems were built directly from the learning samples; they were also applied to the forest type mapping, and results were compared to the other individual AI models.

The study found that the three individual AI models were capable classifiers for this complicated predictive forest type mapping using multisource data. They avoided some significant drawbacks of parametric-based models. The Decision Tree has achieved an overall accuracy of 46.5%, followed by 41.2% of the Artificial Neural Network and 39.0% of the model based on Dempster-Shafer's theory for the 7 forest types.

Although the Decision Tree has achieved the best overall and Kappa accuracies among the three individual AI models, no single model achieved the best user's accuracies and producer's accuracies on all of the 7 forest types. The Z-test has confirmed that there is no significant statistical difference among the three AI models. However, the three classifications appear to be quite different in spatial pattern. This is believed to be because of the different principles the three models are based on. The same reason has also caused them to handle the known data errors differently. Consequently, the Decision Tree was the most adversely affected model in dealing with the sampling error and the error associated with the geology variable. This is because the geology variable was the first variable to be tested in the Decision Tree, and the Decision Tree was very sensitive to the sampling error. The data errors affected the Artificial Neural Network to a lesser extent, but its error propagation mode is more complicated and difficult to interpret. The model based on Dempster-Shafer's theory, however, has been shown to be the most successful in handling data errors. The reason is that it assumes independence among all input variables, so the negative effect of the geology variable was compensated for by other input variables, and it was less sensitive to the sampling error than the Artificial Neural Network and the Decision Tree.

On the other hand, the study has indicated that the combined AI models were better classifiers than the individual AI models for forest type mapping. All of the 34 combined AI models have increased predictive accuracies and improved the visual appearance of the classification maps except one. Several combination methods have been used in the combination processes.

Within the two combined AI models based on the majority voting system, combine1 is the only combined AI model that could not increase the predictive accuracies from those of the Decision Tree. Combine2 has increased overall accuracy at a magnitude of almost 3.1% and Kappa accuracy at a magnitude of over 3.6% from those of the Decision Tree. The combined AI model of D-S1, which is based on Dempster's rule of combination, further increased the predictive accuracies. Among the three combined AI models using simple statistical functions, meanclass was better than medianclass and maxclass in terms of predictive accuracies, but it was not as good as D-S1.

The 18 combined AI models based on fuzzy set theory all have increased the predictive accuracies from those of the Decision Tree to some extent. Their results indicated that among the five measurements of difference, the combined AI models based on bmaxMs achieved consistently higher predictive accuracies than those based on the other four measurements of difference. They also showed that among the four groups of fuzzy membership functions being used, the combined AI models utilizing the fourth group non-linear fuzzy membership function obtained generally better predictive accuracies than those utilizing the other three groups of fuzzy membership functions. Among these combined AI models based on fuzzy set theory, ffamaxclass was the best one of the first stage combination approaches in terms of predictive accuracies. However, the three combined AI models at the second stage combination have further increased the predictive accuracies from those of the first stage combination approaches. The results of vote7 are most encouraging, with an increase in overall accuracy from 46.5% of the Decision Tree to 53.9% and an increase in Kappa accuracy from 37.9% of the Decision Tree to 46.9%. This is significant under the uncertainties and difficulties of forest type mapping. The Z-test has confirmed that most of the combined AI models are significantly different from the Artificial Neural Network and the model based on Dempster-Shafer's theory. More impressive, vote7 is significantly different from the Decision Tree at the 90% confidence level.

Visually comparing the classification maps of the combined AI models and those of the individual AI models has shown that many good features of the three initial classifications have been retained. Some classification maps of the combined AI models appear to be closer to the group truth than the individual AI models. For example, vote7 has perfectly predicted Durras Lake, the power line easement and Willinga River, and it has also predicted Brush Island in large part.

In addition, the study found that several of the combined AI models could handle the known data errors effectively. They are D-S1, the four combined AI models using the measurement of difference of bmaxMs, and the three combined AI models at the second stage combination. It has shown that through some kinds of combination processes, the advantage of the model based on Dempster-Shafer's theory in handling the known data errors could be largely inherited, while the disappointing results of the Decision Tree and the Artificial Neural Network could be largely suppressed.

Furthermore, the combination strategy can also provide prediction confidence measures, which is an attractive feature for the classification problem (Huang & Lees, 2004). This is impossible for a single classifier due to lack of cross-reference. However, the qualitative confidence measure for meanclass (Huang & Lees, 2004) and the quantitative confidence measure for D-S1 (Plate 36) have been estimated by comparing their classification results with those of the three individual AI models. This kind of confidence measure represents a complement to traditional accuracy assessment for the classification problem. Even if they are not very reliable, they can at least provide some additional information to the users and decision makers.

The experiment with Fuzzy Expert Systems found that learning the classification rules directly from samples instead of domain experts is effective and efficient. However, one crucial requirement is to select sufficient and representative learning samples. For example, the failure of fes12 was obviously due to the small size and the inadequate representation of the learning samples derived from the classification map of D-S1. On the contrary, the three Fuzzy Expert Systems generated from large field samples have demonstrated that they were capable classifiers for forest type mapping. Among the three, sarules4 has achieved slightly better predictive accuracies than the Decision Tree. However, it may be over-specified. Pruning the rule base of sarules4 has significantly reduced the size of the rule base. Even though the predictive accuracies were also decreased to some extent, it was worthwhile in terms of generalization. Meanwhile, the Z-test has shown that there is no significant difference among the three Fuzzy Expert Systems. On the other hand, 5%-7% of the area remains unclassified on the three classification maps, but some of these unclassified areas are thought to be reasonable, such as islands and beaches. The three classification maps of the Fuzzy Expert Systems appear to be more fragmented than those of other AI models. The map of certainty factor has demonstrated that the Fuzzy Expert Systems were also able to provide soft classifications.

The most attractive feature of Fuzzy Expert System is its comprehensibility, which represents the classification process as production rules in natural language. Moreover, using fuzzy logic instead of Boolean logic for the inference engine has facilitated the handling of classification uncertainty and partial matching. The explanation machine embodied in the simple and user-friendly user interface could provide answers to user queries of "why", "what if", and "how" questions. However, though these Fuzzy Expert Systems have largely escaped the knowledge acquisition "bottleneck" problem, it was still a very time consuming task to build them. Running these Fuzzy Expert Systems also requires much more time and resources than the other individual and combined AI models. It is believed that this is a trade-off between the efficiency and the comprehensibility.

On the other hand, the comparison of the predictive models of this study and those of previous studies in terms of overall classification accuracy shows that the combination models of this study are better predictive models than those simple AI models used in earlier studies of Fitzgerald, Lees, and Ritman. But they are worse than those advanced AI models used in more recent studies of Gahegan, German, West, and Takatsuka. Nevertheless, the findings do not weaken our conclusions on combining models. Because combining models have several advantages over individual models besides increasing classification accuracy.

In summary, the author believes that the Decision Tree, the Artificial Neural Network, and the model based on Dempster-Shafer's theory are fairly good classifiers for mapping complicated forest types at Anderson *et al.* (1976) level III using multisource data. However, they are still sub-optimal, and none of them is declared to be the best classifier. Because of their different principles, they have advantages and disadvantages of their own in mapping complicated forest types and handling data error. Therefore, relying on any one of these individual models, and implementing it uncritically, is dangerous. Spending time finding a best model and fine-tuning it is not cost effective. More time should be located on understanding the data sets and on studying the application. On the other hand, the author believes that the combination strategy is effective and efficient in mapping complicated forest types. The advantages of the combination strategy include:

- Common advantages of the individual models are retained;
- Good results of each model are combined;
- Conflicting results of each model are either resolved or smoothed;
- Confidence levels of prediction are estimated;
- Classification performance is improved;
- Data error is handled effectively; and
- More time is freed from the modeling process, and can be spent on studying the data and the application.

One pre-requirement, however, is that there have to be at least three good models available in order to implement this strategy.

REFERENCES

Abrahart, R. J., See, L., Kneale, P. E., 1998 New tools for neurohydrologists: using 'network pruning' and 'model breeding' algorithms to discover optimum inputs and architectures, in *Proceedings of the 3rd International Conference on GeoComputation*, 17-19 September 1998, Bristol, U.K.

Abuelgasim, A. A., Gopal, S., Irons, J. R., Strahler, A. H., 1996 Classification of ASAS Multiangle and Multispectral Measurements Using Artificial Neural Networks, *Remote Sensing of Environment*, 57:79-87.

Adamo, J. M., 1980 Fuzzy Decision Trees, *Fuzzy Sets and Systems*, 4, 207-219.

Afonso, P. H., Gomes, M. R., Abrantes, M. G., 1998 A Geomarketing Decision Support System Based on Fuzzy Set Theory, in *Proceedings of the 3rd International Conference on GeoComputation*, 17-19 September 1998, Bristol, U.K.

Agumya, A., and Hunter, G. J., 2002 Responding to the consequences of uncertainty in geographical data, *International Journal of Geographical Information Science*, Vol 16, 405-417.

Albrecht, J., and Guesgen, H. W., 1998 A Combinatorial Fuzzy Set-Theoretic Approach to the Mapping Between Quantitative and Qualitative Data, in *Proceedings of the 3rd International Conference on GeoComputation*, 17-19 September 1998, Bristol, U.K.

Aleksander, I., and Morton, H., 1990 *An Introduction to Neural Computing,* Chapman and all, London.

Alpsan, D., Towsey, M., Ozdamar, O., Tsoi, A. C., Ghista, D. N., 1995 Efficacy of Modified Backpropagation and Optimisation Methods on a Real-world Medical Problem, *Neural Networks*, Vol 8, 945-962.

Altman, D., 1994 Fuzzy set theoretic approaches for handling imprecision in spatial analysis, *International Journal of Geographical Information Science*, Vol 8, 271-289.

Anderson, J. R., Hardy, E. E., Roach, J. T., Witmer, R. E., 1976 *A Land Use and Land Cover Classification System for Use with Remote Sensor Data,* USGS Professional Paper 964, 28p.

Arbia, G., Griffith, D., Haining, R., 1998 Error propagation modelling in raster GIS: overlay operations, *International Journal of Geographical Information Sciences*, Vol 12, 145-167.

Argentiero, P., Chin, R., Beaudet, P., 1982 An Automated Approach to the Design of Decision Tree Classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 4, 51-57.

Arora, M. K., and Foody, G. M., 1997 Log-linear modelling for the evaluation of the variables affecting the accuracy of probabilistic, fuzzy and neural network classifications, *International Journal of Remote Sensing*, Vol 18, 785-798.

Aspinall, R., 1992 An inductive modelling procedure based on Bayes' theorem for analysis of pattern in spatial data, *International Journal of Geographical Information Systems*, Vol 6, 105-121.

Atkinson, P. M., and Tatnall, A. R. L., 1997 Neural networks in remote sensing, *International Journal of Remote Sensing*, Vol 18, 699-709.

Atkinson, P. M., Cutler, M. E. J., Lewis, H., 1997 Mapping sub-pixel proportional land cover with AVHRR imagery, *International Journal of Remote Sensing*, Vol 18, 917-935.

Austin, M. P., Cummingham, R. B., Good, R. B., 1983 Altitudinal distribution of several eucalypt species in relation to other environmental factors in southern New South Wales, *Australian Journal of Ecology*, 8, 169-180.

Bagis, A., 2003 Determining fuzzy membership functions with tabu search – an application to control, *Fuzzy Sets and Systems*, 139, 209-225.

Banai, R., 1993 Fuzziness in Geographical Information Systems: contributions from the analytic hierarchy process, *International Journal of Geographical Information Science*, Vol 7, 315-329.

Bardossy, A., and Samaniego, L., 2002 Fuzzy Rule-Based Classification of Remotely Sensed Imagery, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 40, 362-374.

Barnard, E., and Casasent, D., 1989 A Comparison between Criterion Functions for Linear Classifiers, with an Application to Neural Nets, *IEEE Transactions on Systems, Man and Cybernetics*, Vol 19, 1030-1040.

Barnett, J. A., 1981 Computational methods for a mathematical theory of evidence, in *Proceedings IJCAI-81*, Vancouver, 865-875.

Basden, A., 1983 On the application of expert systems, *International Journal of Man-Machine Studies*, 19, 461-477.

Benediktsson, J. A., and Sveinsson, J. R., 1997 Feature extraction for multisource data classification with artificial neural networks, *International Journal of Remote Sensing*, Vol 18, 727-740.

Benediktsson, J. A., Swain, P. H., Ersoy, O. K., 1990 Neural Network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-28, No 4, 540-551.

Benediktsson, J. A., Swain, P. H., Ersoy, O. K., 1993 Conjugate-gradient neural networks in classification of multisource and very high dimensional remote sensing data, *International Journal of Remote Sensing*, Vol 14, No 5, 2883-2903.

Bezdek, J. C., Ehrlich, R., Full, W., 1984 FCM: The Fuzzy *c*-Means Clustering Algorithm, *Computers and Geosciences*, Vol 10, 191-203.

Bischof, H., Schneider, W., Pinz, A. J., 1992 Multispectral Classification of Landsat-Images using Neural Network, *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-30, No 3, 482-490.

Bishop, Y. S., Fienberg, S. E., and Holland, P. W., 1975 *Discrete Multivariate Analysis—Theory and Practice*, MIT Press, Cambridge, Massachusetts.

Blamire, P. A., 1996 The influence of relative sample size in training artificial neural networks, *International Journal of Remote Sensing*, Vol 17, 223-230.

Bogler, P. L., 1987 Shafer-Dempster reasoning with applications to multisensor Target Identification Systems, *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-17, No 6, 968-977.

Bolstad, P. V., and Lillesand, T. M., 1992 Rule-Based Classification Models: Flexible Integration of Satellite Imagery and Thematic Spatial Data, *Photogrammetric Engineering & Remote Sensing*, Vol 58, 965-971.

Bonde, A., 2000 Fuzzy Expert Systems, http://www.austinlinks.com/Fuzzy/expert-systems.html.

Breiman, L., 1996 Stacked Regression, *Machine Learning*, 24, 49-64.

Breiman, L. H., Friedman, R., Olshen, A., Stone, C. J., 1984 *Classification and regression Trees*, Wadsworth International, Belmont, California, 358 pp.

Briem, G. J., Benediktsson, J. A., Sveinsson, J. R., 2002 Multiple Classifiers Applied to Multisource Remote Sensing Data, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 40, 2291-2299.

Brodley, C. E., and Utgoff, R. E., 1995 Multivariate Decision Trees, *Machine Learning*, 19, 45-77.

Buntine, W., and Niblett, T., 1992 A Further Comparison of Splitting Rules for Decision-Tree Induction, *Machine Learning*, 8, 75-85.

Burrough, P. A., 1989 Fuzzy mathematical methods for soil survey and land evaluation, *Journal of Soil Science*, 40, 477-492.

Burrough, P. A., MacMillan, R. A., van Deursen, W., 1992 Fuzzy classification methods for determining land suitability from soil profile observations and topography, *Journal of Soil Science*, 43, 193-210.

Burrough, P. A., van Gaans, P. F. M., Hootsmans, R., 1997 Continuous classification in soil survey: spatial correlation, confusion and boundaries, *Geoderma*, 77:115-135.

Burrough, P. A., Van Gaans, P. F. M., MacMillan, R. A., 2000 High-resolution landform classification using fuzzy *k*-means, *Fuzzy Sets and Systems*, 113:37-52.

Cannon, R. L., Dave, J. V., Bezdek, J. C., 1986a Efficient Implementation of the Fuzzy *c*-Means Clustering Algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 18, 248-255.

Cannon, R. L., Dave, J. V., Bezdek, J. C., Trivedi, M. M., 1986b Segmentation of a Thematic Mapper Image Using the Fuzzy *c*-Means Clustering Algorithm, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 24, 400-408.

Canters, F., 1997 Evaluating the Uncertainty of Area Estimates Derived from Fuzzy Land-Cover Classification, *Photogrammetric Engineering & Remote Sensing*, Vol 63, 403-414.

Carpenter, G. A., Gjaja, M. N., Gopal, S., Woodcock, C. E., 1997 ART Neural Networks for Remote Sensing: Vegetation Classification from Landsat TM and Terrain Data, *IEEE Transactions on GeoScience and Remote Sensing*, Vol 35, 308-325.

Carpenter, G. A., Gopal, S., Macomber, S., Martens, S., Woodcock, C. E., 1999 A Neural Network Method for Mixture Estimation for Vegetation Mapping, *Remote Sensing of Environment*, 70:138-152.

Carpenter, G. A., Gopal, S., Macomber, S., Martens, S., Woodcock, C. E., and Franklin, J., 1999 A Neural Network Method for Efficient Vegetation Mapping, *Remote Sensing of Environment*, 70:326-338.

Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., Rosen, D. B., 1992 Fuzzy ARCMAP: A Neural Network Architecture for Incremental Supervised Learning of Anolog Multidimentional Maps, *IEEE Transactions on Neural Network*, Vol 3, 698-713.

Carpenter, G. A., 1989 Neural Network Models for Pattern Recognition and Associative Memory, *Neural Networks*, Vol 2, 243-257.

Casey, R. G., and Nagy, G., 1984 Decision Tree Decision Using a Probabilistic Model, *IEEE Transactions on Information Theory*, IT-30, 93-99.

Castro, J. L., Castro-Schez, J. J., Zurita, J. M., 2001 Use of a fuzzy machine learning technique in the knowledge acquisition process, *Fuzzy Sets and Systems*, 123, 307-320.

Cendrowska, J., 1987 PRISM: An algorithm for inducing modular rules, *International Journal of Man-Machine Studies*, 27, 349-370.

Chandrasekaran, B., and Goel, A., 1988 From Numbers to Symbols to Knowledge Structures: Artificial Intelligence Perspectives on the Classification Task, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol 18, 415-424.

Chang, R. L. P., and Pavlidis, T., 1977 Fuzzy Decision Tree Algorithms, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-7, 28-35.

Chen, K. S., Tzeng, Y. C., Chen, C. F., Kao, W. L., 1995 Land-Cover Classification of Multispectral Imagery Using a Dynamic Learning Neural Network, *Photogrammetric Engineering & Remote Sensing*, Vol 61, 403-408.

Chen, K. S., Yen, S. K., Tsay, D. W., 1997 Neural classification of SPOT imagery through integration of intensity and fractal information, *International Journal of Remote Sensing*, Vol 18, 763-783.

Chen, S. M., Ke, J. S., Chang, J. F., 1990 Knowledge Representation Using Fuzzy Petri Nets, *IEEE Transactions on Knowledge and Data Engineering*, Vol 2, 311-319.

Cheng, T., Molenaar, M., Lin, H., 2001 Formalizing fuzzy objects from uncertain classification results, *International Journal of Geographical Information Science*, Vol 15, 27-42.

Civco, D. L., 1993 Artificial neural networks for land-cover classification and mapping, *International Journal of Geographical Information Systems*, Vol 7, No 2, 173-186.

Clark, L. A., and Pregibon, D., 1992 Tree-Based Models, in *Statistical Models In S* (edited by Chambers, J. M., and Hastie, T. J.), Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California, chapter 9.

Clark, P., and Niblett, T., 1989 The CN2 induction algorithm, *Machine Learning*, 3, 261-283.

Cobb, M. A., Petry, F. E., Shaw, K. B., 2000 Fuzzy spatial relationship refinements based on minimum bounding rectangle variations, *Fuzzy Sets and Systems*, 113:111-120.

Cohen, J., 1960 A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement*, Vol XX, 37-46.

Congalton, R. G., 1991 A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data, *Remote Sensing Environment,* 37, 35-46.

Congalton, R. G., and Green, K., 1993 A Practical Look at the Source of Confusion in Error Matrix Generation, *Photogrammetric Engineering and Remote Sensing*, Vol 59, 641-644.

Congalton, R. G., Oderwald, R. G., Mead, R. A., 1983 Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques, *Photogrammetric Engineering and Remote Sensing*, Vol 49, 1671-1678.

Corne, S. A., Carver, S. J., Kunin, W. E., Lennon, J. J., Hees, W. W. S. V., 2000 Using artificial neural network methods to predict forest characteristics in southeast Alaska, in *4th International Conference on Integrating GIS and Environmental Modeling*, September 2000, Banff, Alberta, Canada.

Crevier, D., 1993 *AI: the Tumultuous History of the Search for Artificial Intelligence*, BasicBooks, USA.

Crosetto, M., and Tarantola, S., 2001 Uncertainty and sensitivity analysis: tools for GIS-based model implementation, *International Journal of Geographical Information Science*, Vol 15, 415-437.

Cross, V., and Firat, A., 2000 Fuzzy objects for geographical information systems, *Fuzzy Sets and Systems*, 113:19-36.

Dattatreya, G. R., and Sarma, V. V. S., 1981 Bayesian and Decision Tree Approaches for Pattern Recognition Including Feature Measurement Costs, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3, 293-298.

Davis, J. R., Hoare, J. R. L., Nanninga, P. M., 1986 Developing a Fire Management Expert System for Kakadu National Park, Australia, *Journal of Environmental Management*, 22, 215-227.

Davis, R., Buchanan, B., Shortliffe, E., 1977 Production Rules as a Representation for a Knowledge-Based Consultation Program, *Artificial Intelligence*, 8, 15-45.

Davis, T. J., and Keller, C. P., 1997 Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction, *International Journal of Geographical Information Science*, Vol 11, 409-434.

De Genst, W., Canters, F., Gulinck, H., 2001 Uncertainty Modeling in Buffer Operations Applied to Connectivity Analysis, *Transactions in GIS*, Vol 5, 305-326.

Dietterich, T. G., 1997 Machine learning directions, *AAAI AI Magazine*, 97-136, winter.

Dobermann, A., and Oberthur, T., 1997 Fuzzy mapping of soil fertility - a case study on irrigated Riceland in the Philippines, *Geoderma*, 77:317-339.

Dougherty, M. S., 1997 Meta-analysis of various short term load forecasts, in *Proceedings of Engineering Applications of Neural Networks Conference*, Stockholm, June 1997.

Dragicevic, S., and Marceau, D. J., 2000a A fuzzy set approach for modelling time in GIS, *International Journal of Geographical Information Science*, Vol 14, 225-245.

Dragicevic, S., and Marceau, D. J., 2000b An application of fuzzy logic reasoning for GIS temporal modeling of dynamic process, *Fuzzy Sets and Systems*, 113:69-80.

Dreyer, P., 1993 Classification of Land Cover Using Optimized Neural Nets on SPOT Data, *Photogrammetric Engineering & Remote Sensing*, Vol 59, 617-621.

Drummond, J., 1987 A Framework for Handling Error in Geographic Data Manipulation, *ITC Journal* 1987-1, 73-82.

Dubois, D., and Prade, H., 1990 Consonant Approximation of Belief Functions, *International Journal of Approximate Reasoning*, 4: 419-449.

Dubois, D., Prade, H., Testemale, C., 1993 Weighted Fuzzy Pattern Matching, in *Readings in Fuzzy Sets for Intelligent Systems*, Dubois, D., Prade, H., Yager, R. R., (editors), San Mateo, Calif, Morgan Kaufman Publishers.

Duda, R., Gaschnig, H., Hart, P., 1979 Model Design in PROSPECTOR Consultant System for Mineral Exploration, in *Expert Systems in the Micro-electronic Age*, Michie, D. (Editor), Edinburgh University Press.

Dutta, S., 1991 Approximate Spatial Reasoning: Integrated Qualitative and Quantitative Constraints, *International Journal of Approximate Reasoning*, 5:307-331.

Enbutsu, I., Baba, K., Hara, N., Waseda, K., Nogita, S., 1993 Integration of Multi AI Paradigms for Intelligent Operation Support Systems – Fuzzy Rule Extraction from a Neural Network, *Water Science Technology*, Vol 28, 333-340.

Farreny, H., and Prade, H., 1986 Default and Inexact Reasoning with Possibility Degrees, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-16, 271-276.

Fischer, M. M., and Staufer, P., 1999 Optimization in an Error Backpropagation Neural Network Environment with a Performance Test on a Spectral Pattern Classification Problem, *Geographical Analysis*, Vol 31, 89-108.

Fischer, M. M., Reismann, M., Schindler, K. H., 1999 Parameter estimation in neural spatial interaction modeling by a derivative free global optimization method, in *Proceedings of 4th International Conference on GeoComputation,* 25-28 July 1999, Fredericksburg, Virginia, USA.

Fisher, P., 2000 Sorites paradox and vague geographies, *Fuzzy Sets and Systems*, 113:7-18.

Fisher, P. F., and Pathirana, S., 1990 The Evaluation of Fuzzy Membership of Land Cover Classes in the Suburban Zone, *Remote Sensing of Environment*, 34:121-132.

Fisher, P. F., Mackaness, W. A., Peacegood, G., Wilkinson, G. G., 1988 Artificial intelligence and expert systems in geodata processing, *Progress in Physical Geography*, 12, 371-388.

Fitzgerald, R. W., and Lees, B. G., 1996 Temporal Context in Floristic Classification, *Computers & Geoscience*, Vol 22, No 9, 981-994.

Fitzgerald, R. W., 1994 *The ANU KIOLOA (1994) Remote Sensing & GIS dataset*, distribution documentation, Department of Geography, Australian National University.

Fitzgerald, R. W., and Lees, B. G., 1996 Temporal Context in Floristic Classification, *Computers Remote Sensing of Environment Geoscience*, Vol 22, No 9, 981-994.

Fitzgerald, R. W., and Lees, B. G., 1992 The Application of Neural Networks to the Floristic Classification of Remote Sensing and GIS data in Complex Terrain, in *Proceedings of the 17th ISPRS*, Washington DC, USA, 1992, Vol XXIX, Part B7, 570-573.

Fitzgerald, R. W., and Lees, B. G., 1994 Spatial Context and Scale Relationships in Raster Data for Thematic Mapping in Natural Systems, in *Spatial Data Handling Conference*, Edinburgh, Scotland, Sept 1994.

Foody, G. M., 1994 Ordinal-Level Classification of Sub-Pixel Tropical Forest Cover, *Photogrammetric Engineering & Remote Sensing*, Vol 60, 61-65.

Foody, G. M., 1996 Relating the Land-Cover Composition of Mixed Pixels to Artificial Neural Network Classification Output, *Photogrammetric Engineering & Remote Sensing*, Vol 62, 491-499.

Foody, G. M., 1999 The Continuum of Classification Fuzziness in Thematic Mapping, *Photogrammetric Engineering & Remote Sensing*, Vol 65, 443-451.

Foody, G. M., McCulloch, M. B., Yates, W. B., 1995 Classification of Remotely Sensed Data by an Artificial Neural Network: Issues Related to Training Data Characteristics, *Photogrammetric Engineering & Remote Sensing*, Vol 61, 391-401.

Foody, G. M., 1992 A Fuzzy Sets Approach to the Representation of Vegetation Continua from Remotely Sensed Data: An Example from Lowland Heath, *Photogrammetric Engineering and Remote Sensing*, Vol 58, 221-225.

Foody, G. M., 1995 Land cover classification by an artificial neural network with ancillary information, *International Journal of Geographical Information Science*, Vol 9, 527-542.

Foody, G. M., and Arora, M. K., 1997 An evaluation of some factors affecting the accuracy of classification by an artificial neural network, *International Journal of Remote Sensing*, Vol 18, 799-810.

Foody, G. M., and Boyd, D. S., 1998 Mapping tropical forest biophysical properties from coarse spatial resolution satellite sensor data: applications of neural network and data fusion, in *Proceedings of the 3rd International Conference on GeoComputation*, 17-19 September 1998, Bristol, U.K.

Foody, G. M., Campbell, N. A., Trodd, N. M., Wood, T. F., 1992 Derivation and Applications of Probabilistic Measures of Class Membership from the Maximum-Likelihood Classification, *Photogrammetric Engineering & Remote Sensing*, Vol 58, 1335-1341.

Foschi, P. G., and Smith, D. K., 1997 Detecting Subpixel Woody Vegetation in Digital Imagery Using Two Artificial Intelligence Approaches, *Photogrammetric Engineering & Remote Sensing*, Vol 63, 493-500.

Franssen, H. J. W. M. H., van Eijnsbergen, A. C., Stein, A., 1997 Use of spatial prediction techniques and fuzzy classification for mapping soil pollutants, *Geoderma*, 77:243-262.

Friedl, M. A., and Brodley, C. E., 1997 Decision Tree Classification of Land Cover from Remotely Sensed Data, *Remote Sensing of Environment*, 61:399-409.

Friedman, J. H., 1977 A Recursive Partitioning Decision Rule for Nonparametric Classification, *IEEE Transactions on Computers*, c-26, 404-408.

Gahegan, M., and Takatsuka, M., 1999 Dataspaces as an Organisational Concept for the Neural Classification of Geographic Datasets, in *Proceedings of the 4th International Conference on GeoComputation*, 25-28 July 1999, Fredericksburg, Virginia, USA.

Gahegan, M., and West, G., 1998 The Classification of Complex Geographic Datasets: an Operational Comparison of Artificial Neural Network and Decision Tree Classifiers, in *Proceedings of the 3rd International Conference on GeoComputation*, 17-19 September 1998, Bristol, U.K.

Gahegan, M., 2003 Is inductive machine learning just another wild goose (or might it lay the golden egg)? *International Journal of Geographical Information Science*, Vol 17, 69-92.

Gaines, B. R., and Shaw, M. L. G., 1985 From Fuzzy Logic to Expert Systems, *Information Sciences*, 36, 5-16.

Gaines, B. G., 1993 Fuzzy and Probability Uncertainty Logics, in *Readings in Fuzzy Sets for Intelligent Systems*, Dubois, D., Prade, H., Yager, R. R., (editors), San Mateo, Calif, Morgan Kaufman Publishers.

Gale, S., 1972 Inexactness, Fuzzy Sets, and the Foundations of Behavioral Geography, *Geographical Analysis*, 25, 337-349.

Geman, S., and Geman, D., 1984 Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions in Pattern Analysis and Machine Intelligence*, Vol 6, 721-741.

German, G. W. H., and Gahegan, M. N., 1996 Neural Network Architecture for the Classification of Temporal Image Sequences, *Computers & Geosciences*, Vol 22, 969-979.

German, G., 1999 Neural Network Classifiers for GIS data: improved Search Strategies, in *Proceedings of the 4th International Conference on GeoComputation*, 25-28 July 1999, Fredericksburg, Virginia, USA.

German, G., Gahegan, M., West, G., 1997 Predictive assessment of Neural Network classifiers for applications in GIS, in *Proceedings of 2nd International Conference on GeoComputation*, 26–29 August 1997, University of Otago, Dunedin, New Zealand.

Giarratano, J., and Riley, G., 1998 *Expert Systems Principles and Programming*, PWS Publishing Company.

Gillison, A. N., and Anderson, D. J., 1981 *Vegetation classification in Australia*, Australian National University Press, Canberra.

Gong, P., 1996 Integrated Analysis of Spatial Data from Multiple Sources: Using Evidential Reasoning and Artificial Neural Network Techniques for Geological Mapping, *Photogrammetric Engineering & Remote Sensing*, Vol 62, 513-523.

Gong, P., Pu, R., Chen, J., 1996 Mapping Ecological Land Systems and Classification Uncertainties from Digital Elevation and Forest-Cover Data Using Neural Networks, *Photogrammetric Engineering & Remote Sensing*, Vol 62, 1249-1260.

Goodchild, M. F., 1989 Modeling error in objects and fields, in *Accuracy of Spatial Databases*, Goodchild, M., and Gopal, S. (editors), Taylor & Franeis, New York, Chapter 10.

Goodchild, M. F., Sun, G. Q., Yang, S. R., 1992 Development and test of an error model for categorical data, *International Journal of Geographical Information Systems*, Vol 6, 87-104.

Goodenough, D. G., Goldberg, M., Plunkett, G., Zelek, J., 1987 An Expert System for Remote Sensing, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 25, 349-359.

Goodman, R. M., and Smyth, R., 1988 Decision Tree Design from a Communication Theory Standpoint, *IEEE Transactions on Information Theory*, Vol 34, 979-993.

Goodman, R. M., Higgins, C. M., Miller, J. W., Smyth, P., 1992 Rule-Based Neural Networks for Classification and Probability Estimation, *Neural Computation*, 4, 781-804.

Goodman, R. M., Smyth, P., 1990 Decision Tree design using information theory, *Knowledge Acquisition*, 2, 1-19.

Gopal, S., and Woodcock, C., 1994 Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets, *Photogrammetric Engineering & Remote Sensing*, Vol 60, 181-188.

Gopal, S., Woodcock, C. E., Strahler, A. H., 1999 Fuzzy Neural Network Classification of Global Land Cover from a 1o AVHRR Data Set, *Remote Sensing of Environment*, 67:230-243.

Gordon, J., and Shortliffe, E. H., 1985 A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence*, 26, 323-357.

Gostin, V. A., 1969 *Stratigraphy and Sedimentology of the lower Permian sequence in the Durras-Ulladulla area Sydney Basin, N.S.W.*, PhD thesis, Department of Geology, Australian National University, Canberra, Australia.

Gruijter, J. J. D., Walvoort, D. J. J., van Gaans, P. F. M., 1997 Continuous soil maps - a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models, *Geoderma*, 77:169-195.

Gu, Y. X., Wang, Q. R., Suen, C. Y., 1983 Application of a Multilayer Decision Tree in Computer Recognition of Chinese Characters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5, 83-89.

Guesgen, H. W., and Albrecht, J., 2000 Imprecise reasoning in geographic information systems, *Fuzzy Sets and Systems*, 113:121-131.

Hagen, A., 2003 Fuzzy set approach to assessing similarity of categorical maps, *International Journal of Geographical Information Science*, Vol 17, 235-249.

Haining, R., and Arbia, G., 1993 Error Propagation Through Map Operations, *Technometrics*, Vol 35, 293-3-5.

Hall, G. B., Wang, F., Subaryono, 1992 Comparison of Boolean and fuzzy classification methods in land suitability analysis by using geographical information systems, *Environment and Planning*, 24, 497-516.

Hansen, M., Dubayah, R., Defries, R., 1996 Classification trees: an alternative to traditional land cover classifiers, *International Journal of Remote Sensing*, Vol 17, 1075-1081.

Hara, Y., Atkins, R. G., Yueh, S. H., Shin, R. T., Kong, J. A., 1994 Application of Neural Networks to Radar Image Classification, *IEEE Transactions on GeoScience and Remote Sensing*, Vol 32, 100-109.

Hart, P. E., Duda, R. O., Einaudi, M. T., 1978 PROSPECTOR-A Computer-Based Consultation System for Mineral Exploration, *Mathematical Geology*, Vol 10, 589-610.

Hart, E. B., 1985 Experience in the use of an inductive system in knowledge engineering, in *Research and Development in Expert Systems*, Bramer, M. A. (Editor), Cambridge, Cambridge University Press.

Hartmann, C. P., Varshney, P. K., Mehrotra, K. G., Gerberich, C. L., 1982 Application of Information Theory to the Construction of Efficient Decision Trees, *IEEE Transactions on Information Theory*, IT-28, 565-577.

Heermann, P. D., and Khazenie, N., 1992 Classification of multispectral remote sensing data using a back-propagation neural network, *IEEE Transactions on GeoScience and Remote Sensing*, Vol 30, 81-88.

Hepner. F., Logan, T., Ritter, N., Bryant, N., 1990 Artificial Neural Network Classification Using a Minimal Training Set: Comparison to Conventional Supervised Classification, *Photogrammetric Engineering and Remote Sensing*, Vol 56, No 4, 469-473.

Heuvelink, G. B. M., and Burrough, P. A., 1993 Error propagation in cartographic modelling using Boolean logic and continuous classification, *International Journal of Geographical Information Systems*, Vol 7, 231-246.

Heuvelink, G. B. M., Burrough, P. A., Stein, A., 1989 Propagation of errors in spatial modelling with GIS, *International Journal of Geographical Information Systems*, Vol 3, 303-322.

Hornik, K., Stinchcombe, M., White, H., 1989 Multilayer feedforward networks are universal approximators, *Neural Networks*, vol 2, 359-366.

Huang, X., and Jensen, J. R., 1997 A Machine-Learning Approach to Automated Knowledge-Base Building for Remote Sensing Image Analysis with GIS Data, *Photogrammetric Engineering and Remote Sensing*, Vol 63, 1185-1194.

Huang, Z., and Lees, B. G., 2004, Combining Non-Parametric Models for Multisource Predictive Forest Mapping, *Photogrammetric Engineering and Remote Sensing*, Vol 70, 415-426.

Hutchinson, C. F., 1982 Techniques for Combining Landsat and Ancillary Data for Digital Classification Improvement, *Photogrammetric Engineering and Remote Sensing,* Vol 48, 123-130.

Irvin, B. J., Ventura, S. J., Slater, B. K., 1997 Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin, *Geoderma*, 77:137-154.

Itami, R. M., MacLaren, G., Hirst, K., 2000 Integrating the Analytical Hierarchy Process with GIS to capture expert knowledge for Land Capability Assessment, in *4th International Conference on Integrating GIS and Environmental Modeling*, September 2000, Banff, Alberta, Canada.

Ito, Y., Omatu, S., 1999 Extended LVQ Neural Network approach to land cover mapping, *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-37, No 1, 313-317.

Janikow, C. Z., 1998 Fuzzy Decision Trees: Issues and Methods, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 28, Issue 1, pp 1-14.

Jarvis, C. H., and Stuart, N., 1996 The Sensitivity of a Neural Network for Classifying Remotely Sensed Imagery, *Computers & Geosciences*, Vol 22, 959-967.

Jiang, H., and Eastman, J. R., 2000 Application of fuzzy measures in multi-criteria evaluation in GIS, *International Journal of Geographical Information Science*, Vol 24, 173-184.

Kaminsky, E. J., Barad, H., Brown, W., 1997 Textural neural network and version space classifiers for remote sensing, *International Journal of Remote Sensing*, Vol 18, 741-762.

Kanellopoulos, I., and Wilkinson, G. G., 1997 Strategies and best practice for neural network image classification, *International Journal of Remote Sensing*, Vol 18, 711-725.

Kent, J. T., and Mardia, K. V., 1988 Spatial Classification Using Fuzzy Membership Models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 10, 659-670.

Key, J., Maslanik, J. A., Schweiger, A. J., 1989 Classification of Merged AVHRR and SMMR Arctic Data with Neural Networks, *Photogrammetric Engineering and Remote Sensing*, Vol 55, 1331-1338.

Kollias, V. J., and Voliotis, A., 1991 Fuzzy reasoning in the development of geographical information systems FRSIS: a prototype soil information system with fuzzy retrieval capabilities, *International Journal of Geographical Information Science*, Vol 5, 209-223.

Kononenko, I., Bratko, I., Roskar, E., 1984 *Experiments in automatic learning of medical diagnostic rules* (Technical Report) Ljubljana, Yugoslavia: E.Kardelj University, Faculty of Electrical Engineering.

Kontoes, C., Wilkinson, G. C., Burrill, A., Goffredo, S., Megier, J., 1993 An experimental system for the integration of GIS data in knowledge-based images analysis for remote sensing of agriculture, *International Journal of Geographical Information Systems*, Vol 7, 247-262.

Kosko, B., 1990 Fuzziness Vs Probability, *International Journal of General Systems*, Vol 17, 211-240.

Kurzynski, M. W., 1983 The Optimal Strategy of a Tree Classifier, *Pattern Recognition*, Vol 16, 81-87.

Kyburg, H. E., 1987 Bayesian and Non-Bayesian Evidential Updating, *Artificial Intelligence*, 31, 271-293.

Laba, M., Gregory, S. K., Braden, J., Ogurcak, D., Hill, E., Fegraus, E., Fiore, J., DeGloria, S. D., 2002 Conventional and fuzzy accuracy assessment of the New York Gap Analysis Project land cover map, *Remote Sensing of Environment*, 81, 443-455.

Lacher, R. C., Hruska, S. I., Kunicky, D. C., 1992 Back-Propagation Learning in Expert Networks, *IEEE Transactions on Neural Network*, Vol 3, 62-72.

Laffan, S. W., 1998 Visualising Neural Network Training in Geographic Space, in *Proceedings of the 3rd International Conference on GeoComputation*, 17-19 September 1998, Bristol, U.K.

Lam, D., and Pupp, C., 1996 Integration of GIS, Expert Systems, and Modeling for State-of-Environment Reporting, in *GIS and Environmental Modeling: Progress and Research Issues,* Goodchild, M. F., Steyart, L., Parks, B., Crane, M., Johnston, C., Maidment, D., Glendinning, S., (editors), GIS World Inc, Fort Collins, Colorado, page 419-422.

Landeweerd, G. H., Timmers, T., Gelsema, E. S., Bins, M., Halie, M. R., 1983 Binary Tree Versus Single Level Tree Classification of White Blood Cells, *Pattern Recognition*, Vol 16, 571-577.

Lanter, D. P., and Veregin, H., 1992 Research Paradigm for Propagating Error in Layer-Based GIS, *Photogrammetric Engineering and Remote Sensing*, Vol 58, 825-833.

Leary, M., 1988 What are Expert Systems and how can they contribute to Planning, *Planning Outlook*, Vol 31, 7-12.

LeBlanc, M., and Tibshirani, R., 1996 Combining Estimates in Regression and Classification, *Journal of the American Statistical Association*, Vol 91, 1641-1650.

Lee, J., Snyder, P. K., Fisher, P. F., 1992 Modeling the Effect of Data Errors on Feature Extraction from Digital Elevation Models, *Photogrammetric Engineering and Remote Sensing*, Vol 58, 1461-1467.

Lee, J., Snyder, P. K., Fisher, P. F., 1992 Modeling the Effect of Data Errors on Feature Extraction from Digital Elevation Models, *Photogrammetric Engineering and Remote Sensing*, Vol 58, 1461-1467.

Lee, N. S., Grize, Y. L., Dehnad, K., 1987 Quantitative Models for Reasoning under Uncertainty in Knowledge-Based Expert Systems, *International Journal of Intelligence Systems*, Vol 2, 15-38.

Lee, T., Richards, J. A., Swain, P. H., 1987 Probabilistic and Evidential Approaches for Multisource Data Analysis, *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-25, No 3, 283-292.

Lees, B. G., and Ritman, K., 1991 Decision-Tree and Rule-Induction Approach to Integration of Remotely Sensed and GIS Data in Mapping Vegetation in Disturbed or Hilly Environments, *Environmental Management*, Vol 15, no 6, 823-831.

Lees, B. G., 1996a Improving the Spatial Extension of Point Data, Http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/lees_brian/santafe2.htm.

Lees, B. G., 1996b Neural Network applications in the Geosciences: an introduction, *Computers & Geosciences*, Vol 22, No 9, 955-957.

Lees, B. G., 1996c Sampling Strategies for machine learning using GIS, in *GIS and Environmental Modeling: Progress and Research Issues,* Goodchild, M. F., Steyart, L., Parks, B., Crane, M., Johnston, C., Maidment, D., Glendinning, S., (editors), GIS World Inc, Fort Collins, Colorado, page 39-43.

Lees, B. G., 1996d Species Mapping in Forest Ecosystems, in *Remote Sensing and Computer Technology for Natural Resource Assessment,* Saramaki, J., Koch, B., Lund, H. G. (editors), Albert-Ludwigs-Universitat Freihurg Press.

Lein, J. K., 1993 Applying Expert Systems Technology to Carrying Capacity Assessment: a Demonstration Prototype, *Journal of Environmental Management*, 37, 63-84.

Lesmo, L., Saitta, L., Torasso, P., 1985 Evidence combination in expert systems, *International Journal of Man-Machine Studies*, 22, 307-126.

Lesmo, L., Saitta, L., Torasso, P., 1993 Learning of fuzzy production rules for medical diagnosis, in *Readings in Fuzzy Sets for Intelligent Systems*, Dubois, D., Prade, H., Yager, R. R., (editors), San Mateo, Calif, Morgan Kaufman Publishers.

Leung, Y., 1982 Approximate Characterization of Some Fundamental Concepts of Spatial Analysis, *Geographical Analysis*, Vol 14, 29-40.

Leung, Y., 1987 On the Imprecision of Boundaries, *Geographical Analysis*, Vol 19, 125-150.

Leung, Y., and Leung, K. S., 1993 An Intelligent expert system shell for knowledge-based Geographical Information Systems: 2. Some applications, *International Journal of Geographical Information Science*, Vol 7, 201-213.

Leung, Y., and Leung, K. S., 1993 An Intelligent expert system shell for knowledge-based Geographical Information Systems: 1. The tools, *International Journal of Geographical Information Science*, Vol 7, 189-199.

Li, R. P., Mukaidono, M., Turksen, I. B., 2002 A fuzzy neural network for pattern classification and feature selection, *Fuzzy Sets and Systems*, 130, 101-108.

Li, X., and Dubes, R. C., 1986 Tree Classifier Design with a Permutation Statistic, *Pattern Recognition*, Vol 19, 229-235.

Liepins, G., Goeltz, R., and Rush, R., 1990 Machine Learning Techniques for Natural Resources Data Analysis, *AI Applications*, Vol 4, 9-17.

Lin, C. J., 2004 A GA-based neural fuzzy system for temperature control, *Fuzzy Sets and Systems*, 143, 311-333.

Lindley, D. V., 1987 The Probability Approach to the Treatment of Uncertainty in Artificial Intelligence and Expert Systems, *Statistical Science*, Vol 2, 17-24.

Lippmann, R. P., 1987 An Introduction to Computing with Neural Nets, *IEEE ASSP Magazine*, April, 4-22.

Liu, W. Z., and White, A. P., 1994 The Importance of Attribute Selection Measures in Decision Tree Induction, *Machine Learning*, 15, 25-41.

Liu, Z. K., and Xiao, J. Y., 1991 Classification of remotely sensed image data using artificial neural networks, *International Journal of Remote Sensing*, Vol 12, No 11, 2433-2438.

Lodwick, W. A., 1989 Developing confidence limits on errors of suitability analyses in geographical information systems, in *Accuracy of Spatial Databases*, Goodchild, M., and Gopal, S. (editors), Taylor & Francis, New York, Chapter 6.

Loh, D. K., and Rykiel, E. J., 1992 Integrated Resource Management Systems: Coupling Expert Systems with Data-Base Management and Geographic Information Systems, *Environmental Management*, Vol 16, 167-177.

Lunetta, R. S., Congalton, R. G., Fenstermaker, L. K., Jensen, J. R., McGwire, K. C., Tinney, L. R., 1991 Remote Sensing and Geographic Information System Data Integration: Error Sources and Research Issues, *Photogrammetric Engineering and Remote Sensing*, Vol 57, 677-687.

Ma, Z., and Redmond, R. L., 1995 Tau Coefficients for Accuracy Assessment of Classification of Remote Sensing Data, *Photogrammetric Engineering and Remote Sensing*, Vol 61, 435-439.

MacDougall, E. B., 1975 The accuracy of map overlays, *Landscape Planning*, Vol 2, 23-30.

Mackay, D. S., and Robinson, V. B., 2000 A multiple criteria decision support system for testing integrated environmental models, *Fuzzy Sets and Systems*, 113:53-67.

MacMillan, R. A., Pettapiece, W. W., Nolan, S. C., Goddard, T. W., 2000 A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic, *Fuzzy Sets and Systems*, 113:81-109.

Maiers, J., and Sherif, Y. S., 1985 Applications of Fuzzy Set Theory, *IEEE Transactions on System, Man, and Cybernetics*, Vol 15, 175-189.

Mason, D. C., Corr, D. G., Cross, A., Hogg, D. C., Lawrence, D. H., Petrou, M., and Tailor, A. M., 1988 The use of digital map data in the segmentation and classification of remotely-sensed images, *International Journal of Geographical Information Systems*, Vol 2, 195-215.

McBratney, A. B., and Odeh, I. O. A., 1997 Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions, *Geoderma*, 77:85-113.

McBratney, A. B., and Gruijter, J. J. D., 1992 A continuum approach to soil classification by modified fuzzy *k*-means with extragrades, *Journal of Soil Science*, 43, 159-175.

McCarthy, J., Abrahams, P. W., Edwards, D. J., Hart, T. P., Levin, M. I., 1965 *LISP 1.5 programmer's manual* (2^{nd} ed.), Cambridge, MA:MIT Press.

McIntosh, R. P., 1978 *Phytosociology*, Dowden Hutchinson & Ross, Stroudburg.

McIver, D. K., and Friedl, M. A., 2001 Estimating Pixel-Scale Land Cover Classification Confidence Using Nonparametric Machine Learning Methods, *IEEE Transcations on Geoscience and Remote Sensing*, Vol 39, 1959-1967.

Meisel, W. S., and Michalopoulos, D. A., 1973 A Partitioning Algorithm with Application in Pattern Classification and the Optimization of Decision Trees, *IEEE Transactions on Computers*, c-22, 93-103.

Merz, C. J., 1999 Using Correspondence Analysis to Combine Classifiers, *Machine Learning*, 36, 33-58.

Michaelsen, J., Schimel, D. S., Friedl, M. A., Davis, F. W., Dubayah, R. C., 1994 Regression Tree Analysis of Satellite and terrain data to guide vegetation sampling and surveys, *Journal of Vegetation Science*, 5:673-686.

Mingers, J., 1986 Expert Systems-Experiments with Rule Induction, *Journal of Operational Research Society*, Vol 37, 1031-1037.

Mingers, J., 1987a Expert Systems-Rule Induction with Statistical Data, *Journal of Operational Research Society*, Vol 38, 39-47.

Mingers, J., 1987b Rule Induction with Statistical Data-A Comparison with Multiple Regression, *Journal of Operational Research Society*, Vol 38, 347-351.

Mingers, J., 1989a An Empirical Comparison of Pruning Methods for Decision Tree Induction, *Machine Learning*, 4, 227-243.

Mingers, J., 1989b An empirical comparison of selection measures for Decision-Tree induction, *Machine Learning*, 3, 319-342.

Mojirsheiban, M., 1999 Combining Classifiers via Discretization, *Journal of the American Statistical Association*, Vol 94, 600-609.

Moon, W. M., 1990 Integration of Geophysical and Geological Data Using Evidential Belief Function, *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-28, No 4, 711-720.

Moore, D. M., Lees, B. G., Davey, S. M., 1991 A New Method for Predicting Vegetation Distributions using Decision Tree Analysis in a Geographic Information System, *Environmental Management*, vol 15, no 1, 59-71.

Murnion, S. D., 1996 Comparison of Back Propagation and Binary Diamond Neural Networks in the Classification of a Landsat TM Image, *Computers & Geosciences*, Vol 22, 995-1001.

Murphy, C. K., 2000 Combing belief functions when evidence conflicts, *Decision Support Systems*, 29, 1-9.

Newcomer, J. A., and Szajgin, J., 1984 Accumulation of thematic map errors in digital overlay analysis, *American Cartographer*, Vol 11, 58-62.

Niblett, T., and Bratko, I., 1987 Learning decision rules in noisy domains, in *Research and development in expert systems*, Bramer, M. A., (Editor), Cambridge, Cambridge University Press.

Obermeier, K., Barron, J., 1989 Time to get fired up, *Byte* (August), 217-224.

Olaru, C., and Wehenkel, L., 2003 A complete fuzzy decision tree technique, *Fuzzy Sets and Systems*, 138, 221-254.

Openshaw, S., 1989 Learning to live with errors in spatial databases, in *Accuracy of Spatial Databases*, Goodchild, M., and Gopal, S. (editors), Taylor & Francis, New York, Chapter 23.

Openshaw, S., and Turton, I., 1996 A Parallel Kohonen Algorithm for the Classification of Large Spatial Datasets, *Computers & Geosciences*, Vol 22, 1019-1026.

Openshaw, S., Charlton, M., Carver, S., 1991 Error propagation: a Monte Carlo simulation, in *Handling Geographical Information*, Masser, I., and Blakemore, M. (editors), Harlow: Longman, 78-101.

Orchard, R. A., 1998 *FuzzyCLIPS Version 6.04A User's Guide*, Integrated Reasoning, Institute for Information Technology, National Research Council Canada.

Pal, M., and Mather, P. M., 2003 An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment*, 86, 554-565.

Paola, J. D., Schowengerdt, R. A., 1995 A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification, *IEEE Transactions on Geoscience and Remote Sensing*, Vol GE-33, No 4, 981-996.

Partridge, D., 1996 Representation of Knowledge, in *Artificial Intelligence*, Boden, M. A. (Editor), Academic Press, San Diego, Chapter 3.

Pax-Lenney, M., Woodcock, C. E., Macomber, S. A., Gopal, S., Song, C., 2001 Forest mapping with a generalized classifier and Landsat TM data, *Remote Sensing of Environment*, 77:241-250.

Payne, H. J., and Meisel, W. S., 1977 An Algorithm for Constructing Optimal Binary Decision Trees, *IEEE Transactions on Computers*, c-26, 905-916.

Pearl, J., 1990 Reasoning with Belief Functions: An Analysis of Compatibility, *International Journal of Approximate Reasoning*, 4:363-389.

Peddle, D. R., 1995 Knowledge Formulation for Supervised Evidential Classification, *Photogrammetric Engineering & Remote Sensing*, Vol 61, 409-417.

Peddle, D. R., and Franklin, S. E., 1993 Classification of Permafrost Active Layer Depth from Remotely Sensed and Topographic Evidence, *Remote Sensing of Environment*, 44:67-80.

Peddle, D. R., and Franklin, S. E., 1992 Multisource evidential classification of surface cover and frozen ground, *International Journal of Remote Sensing*, Vol 13, No 17, 3375-3380.

Penaloza, M. A., and Welch, R. M., 1996 Feature Selection for Classification of Polar Regions Using a Fuzzy Expert System, *Remote Sensing of Environment*, 58:81-100.

Petrakos, M., Benediktsson, J. A., Kanellopoulos, I., 2001 The Effect of Classifier Agreement on the Accuracy of the Combined Classifier in Decision Level Fusion, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 39,2539-2546.

Power, C., Simms, A., White, R., 2001 Hierarchical fuzzy pattern matching for the regional comparison of land use maps, *International Journal of Geographical Information Science*, Vol 15, 77-100.

Prade, H., 1985 A Computational Approach to Approximate and Plausible Reasoning with Applications to Expert Systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7, 260-283.

Prisley, S. P., and Smith, J. L., 1987 Using Classification Error Matrices to Improve the Accuracy of Weighted Land-Cover Models, *Photogrammetric Engineering and Remote Sensing*, Vol 53, 1259-1263.

Qing-Yun, S., and Fu, K. S., 1983 A Method for the Design of Binary Tree Classifier, *Pattern Recognition*, Vol 16, 593-603.

Quinlan, J. R., and Rivest, R. L., 1989 Inferring Decision Trees Using the Minimum Description Length Principle, *Information and Computation*, 80, 227-248.

Quinlan, J. R., 1987 Simplifying decision trees, *International Journal of Man-Machine Studies*, 27, 221-234.

Quinlan, J. R., 1990 Decision Trees and Decision making, *IEEE Transactions on Systems, Man and Cybernetics*, Vol 20, 339-346.

Quinlan, J. R., 1986 Induction of Decision Trees, *Machine Learning*, 1, 81-106.

Quinlan, J. R., 1993 *C4.5: Programs for Machine Learning*, Morgan Kauffman.

Ramsey, C. L., Reggia, J. A., Nau, D. S., Ferrentino, A., 1986 A comparative analysis of methods for expert systems, *International Journal of Man-Machine Studies*, 24, 475-499.

Rendell, L., Cho, H., Seshu, R., 1989 Improving the Design of Similarity-Based Rule-Learning Systems, *International Journal of Expert Systems*, Vol 2, 97-133.

Richards, J. A., 1993 *Remote Sensing Digital Image Analysis*, Springer-Verlag, Berlin.

Ripple, W. J., and Ulshoefer, V. S., 1987 Expert Systems and Spatial Data Models for Efficient Geographic Data Handling, *Photogrammetric Engineering & Remote Sensing*, Vol 53, 1431-1433.

Ritter, N. D., and Hepner, G. F., 1990 Application of An Artificial Neural Network to Land-Cover Classification of Thematic Mapper Imagery, *Computers & Geosciences*, Vol 16, 873-888.

Robinson, V. B., Frank, A. U., Karimi, H. A., 1987 Expert Systems for Geographic Information Systems in Resource Management, *AI Applications*, Vol 1, 47-57.

Robinson, V. B., 2000 Individual and multipersonal fuzzy spatial relations acquired using human-machine interaction, *Fuzzy Sets and Systems*, 113:133-145.

Robinson, V. B., and Frank, A. U., 1987 Expert Systems for Geographic Information Systems, *Photogrammetric Engineering and Remote Sensing*, Vol 53, 1435-1441.

Rosenfield, G. H., and Fitzpatrick-Lins, K., 1986 A Coefficient of Agreement as a Measure of Thematic Classification Accuracy, *Photogrammetric Engineering and Remote Sensing*, Vol 52, 223-227.

Rounds, E. M., 1980 A Combined Nonparametric Approach to Feature Selection and Binary Decision Tree Design, *Pattern Recognition*, Vol 12, 313-317.

Roussel, P., 1975 *Prolog: Manuel de Reference et d'Utilisation*, Marseille-Luminy: Groupe d'Intelligence Artificielle.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986 Learning representations by back-propagating errors, *Nature*, Vol 323, 533-535.

Safavian, S. R., and Landgrebe, D., 1991 A Survey of Decision Tree Classifier Methodology, *IEEE Transactions on Systems, Man and Cybernetics*, Vol 21, 660-674.

Salu, Y., and Tilton, J., 1993 Classification of Multispectral Image Data by the Binary Diamond Neural and by Nonparametric, Pixel-by-Pixel Methods, *IEEE Transactions on GeoScience and Remote Sensing*, Vol 31, 606-617.

Schank, R. C., 1988 What is AI, Anyway? *AI Applications*, Vol 2, 12-16.

Schuermann, J., and Doster, W., 1984 A Decision Theoretic Approach to Hierarchical Classifier Design, *Pattern Recognition*, Vol 17, 359-369.

See, L., Abrahart, R. J., Openshaw, S., 1998 An Integrated Neuro-Fuzzy_Statistical Approach to Hydrological Modelling, in *Proceedings of the 3rd International Conference on GeoComputation*, 17-19 September 1998, Bristol, U.K.

Serpico, S. B., and Roli, F., 1995 Classification of Multisensor Remote-Sensing Images by Structured Neural Networks, *IEEE Transactions on GeoScience and Remote Sensing*, Vol 33, 562-577.

Sexton, R. S., and Dorsey, R. E., 2000 Reliable Classification using neural networks: a genetic algorithm and backpropagation comparison, *Decision Support Systems*, 30, 11-22.

Shackelford, A. K., and Davis, C. H., 2003 A Hierarchical Fuzzy Classification Approach for High-Resolution Multispectral Data Over Urban Areas, *IEEE Transactions on GeoScience and Remote Sensing*, Vol 41, 1920-1932.

Shafer, G., 1990 Perspectives on the Theory and Practice of Belief Functions, *International Journal of Approximate Reasoning*, 4:323-362.

Shafer, G., 1976 *A Mathematical Theory of Evidence*, Princeton, New Jersey, Princeton University Press.

Shafer, G., 1987 Probability Judgement in Artificial Intelligence and Expert Systems, *Statistical Science*, Vol 2, 3-16.

Shafer, G., and Logan, R., 1987 Implementing Dempster's rule for hierarchical evidence, *Artificial Intelligence*, 33, 271-298.

Shavlik, J. W., Mooney, R. J., Towell, G. G., 1991 Symbolic and Neural Learning Algorithms: An Experimental Comparison, *Machine Learning*, 6, 111-143.

Shortliffe, E. H., and Buchanan, B. G., 1985 A Model of Inexact Reasoning in Medicine, in *Rule-based Expert Systems*, Addison-Wesley Publication, 233-262.

Skidmore, A. K., 1989 An Expert System Classifies Eucalypt Forest Types Using Thematic Mapper Data and a Digital Terrain Model, *Photogrammetric Engineering and Remote Sensing*, Vol 55, 1449-1464.

Skidmore, A. K., and Turner, B. J., 1988 Forest mapping accuracies are improved using a supervised nonparametric classifier with SPOT data, *Photogrammetric Engineering Remote Sensing of Environment Remote Sensing*, Vol 54, No 10, 1415-1421.

Skidmore, A .K., and Turner, B. J., 1995 Remote sensing and Geographical Information Systems as forestry tools: a critique, in *Proceedings of the Institute of Foresters of Australia 16th Biennial Conference*, Ballarat, Victoria, April, 1995.

Skidmore, A. K., Turner, B. J., Brinkhof, W. B., Knowles, E., 1997 Performance of a Neural Network: Mapping forests using GIS and Remotely Sensed data, *Photogrammetric Engineering Remote Sensing of Environment Remote Sensing*, Vol 63, No 5, 501-514.

Skidmore, A. K., Watford, F., Luckananurug, P., Ryan, P. J., 1996 An Operational GIS Expert System for Mapping Forest Soils, *Photogrammetric Engineering and Remote Sensing*, Vol 62, 501-511.

Skidmore, A. K., Wood, G. B., Shepherd, K. R., 1987 Remotely sensed digital data in forestry: a review, *Australian Forestry*, 50, 40-53.

Smith, T. R., 1984 Artificial Intelligence and its Applicability to Geographical Problem Solving, *Professional Geographer*, Vol 36, 147-158.

Smits, P. C., 2002 Multiple Classifier Systems for Supervised Remote Sensing Image Classification Based on Dynamic Classifier Selection, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 40, 801-813.

Spiegelhalter, D. J., 1987 Probabilistic Expert Systems in Medicine: Practical Issues in Handling Uncertainty, *Statistical Science*, Vol 2, 25-30.

Srinivasan, A., and Richards, J. A., 1990 Knowledge-based techniques for multi-source classification, *International Journal of Remote Sensing*, Vol 11, No 3, 505-525.

Steele, B. M., 2000 Combining Multiple Classifiers: An Application Using Spatial and Remotely Sensed Information for Land Cover Type Mapping, *Remote Sensing of Environment*, 74:545-556.

Stefanakis, E., Vazirgiannis, M., Sellis, T., 1999 Incorporating fuzzy set methodologies in a DBMS repository for the application domain of GIS, *International Journal of Geographical Information Science*, Vol 13, 657-675.

Stockwell, D. R. B., Davey, S. M., Davis, J. R., Noble, I. R., 1990 Using induction of Decision Trees to predict Greater Glider density, *AI Application*, Vol 4, No 4, 33-43.

Story, M., and Congalton, R. G., 1986 Accuracy assessment: A user's perspective, *Photogrammetric Engineering and Remote Sensing*, Vol 52, 397-399.

Strat, T. M., and Lowrance, J. D., 1989 Explaining Evidential Analysis, *International Journal of Approximate Reasoning*, 3:299-353.

Sun, D., Hnatiuk, R. J., Neldner, V. J., 1996 *Vegetation classification and mapping systems for Australian forest management*, Bureau of Resource Science/Australian Nature Conversation Agency, Canberra.

Swain, P. H., and Hauska, H., 1977 The Decision Tree Classifier: Design and Potential, *IEEE Transactions on GeoScience Electronics*, Vol 15, 142-147.

Tatem, A. J., Lewis, H. G., Atkinson, P. M., Nixon, M. S., 2002 Super-resolution land cover pattern prediction using a Hopfield neural network, *Remote Sensing of Environment*, 79, 1-14.

Taylor, J. R., 1982 *An Introduction to Error Analysis*, Mill Valley, CA: University Science Books.

Townsend, P. A., 2000 A Quantitative Fuzzy Approach to Assess Mapped Vegetation Classifications for Ecological Applications, *Remote Sensing of Environment*, 72:253-267.

Trivedi, M. M., and Bezdek, J. C., 1986 Low-Level Segmentation of Aerial Images with Fuzzy Clustering, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-16, 589-597.

Turksen, I. B., 1991 Measurement of membership functions and their acquisition, *Fuzzy Sets and Systems*, 40:5-38.

Tzeng, Y. C., Chen, K. S., Kao, W. L., Fung, A. K., 1994 Dynamic Learning Neural Network for Remote Sensing Applications, *IEEE Transactions on GeoScience and Remote Sensing*, Vol 32, 1096-1102.

Unwin, D. J., 1995 Geographical Information Systems and the Problem of 'error and uncertainty', *Progress in Human Geography*, Vol 19, 549-558.

Urbanski, J. A., 1999 The use of fuzzy sets in the evaluation of the environment of coastal waters, *International Journal of Geographical Information Science*, Vol 13, 723-730.

Utgoff, P. E., 1989 Incremental Induction of Decision Trees, *Machine Learning*, 4, 161-186.

Veregin, H., 1989 Error modelling for the map overlay operation, in *Accuracy of Spatial Databases*, Goodchild, M., and Gopal, S. (editors), Taylor & Franeis, New York, Chapter 1.

Veregin, H., 1994 Integration of Simulation Modelling and Error Propagation for the Buffer Operation in GIS, *Photogrammetric Engineering and Remote Sensing*, Vol 60, 427-435.

Veregin, H., 1995 Developing and testing of an error propagation model for GIS overlay operations, *International Journal of Geographical Information Systems*, Vol 9, 595-619.

Veregin, H., 1996 Error Propagation through the Buffer Operation for Probability Surface, *Photogrammetric Engineering and Remote Sensing*, Vol 62, 419-428.

Voorbraak, F., 1989 A Computationally efficient approximation of Dempster-Shafer theory, *International Journal of Man-Machine Studies*, 30, 525-536.

Voorbraak, F., 1991 On the justification of Dempster's rule of combination, *Artificial Intelligence*, 48:171-197.

Walley, P., 1996 Measures of uncertainty in expert systems, *Artificial Intelligence*, 83, 1-58.

Walsh, S. J., Lightfoot, D. R., Bulter, D. R., 1987 Recognition and Assessment of Error in Geographic Information Systems, *Photogrammetric Engineering and Remote Sensing*, Vol 53, 1423-1430.

Wang, L. X., and Mendel, J. M., 1992 Generating Fuzzy Rules by Learning from Examples, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol 22, 1414-1427.

Wang, Q. R., and Suen, C. Y., 1987 Large Tree Classifier with Heuristic Search and Global Training, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9, 91-101.

Wang, F., 1990a Fuzzy Supervised Classification of Remote Sensing Images, *IEEE Transactions on Geoscience and Remote Sensing*, Vol 28, 194-201.

Wang, F., 1990b Improving Remote Sensing Image Analysis through Fuzzy Information Representation, *Photogrammetric Engineering and Remote Sensing*, Vol 56, 1163-1169.

Wang, F., 1994 Towards a natural language user interface: an approach of fuzzy query, *International Journal of Geographical Information Science*, Vol 8, 143-162.

Wang, F., 2000 A fuzzy grammar and possibility theory-based natural language user interface for spatial queries, *Fuzzy Sets and Systems*, 113:147-159.

Wang, F., and Hall, B., 1996 Fuzzy representation of geographical boundaries in GIS, *International Journal of Geographical Information Science*, Vol 10, 573-590.

Wang, F., Hall, G. B., Subaryono, 1990 Fuzzy information representation and processing in conventional GIS software: database design and application, *International Journal of Geographical Information Science*, Vol 4, 261-283.

Warner, T. A., and Shank, M., 1997 An Evaluation of the Potential for Fuzzy Classification of Multispectral Data Using Artificial Neural Networks, *Photogrammetric Engineering & Remote Sensing*, Vol 63, 1285-1294.

Whalen, T., and Schott, B., 1983 Issues in fuzzy production systems, *International Journal of Man-Machine Studies*, 19, 57-71.

Wharton, S. W., 1987 *IEEE Transactions on GeoScience and Remote Sensing*, 25, 272-282.

Wilkinson, G. G., and Megier, J., 1990 Evidential reasoning in a pixel classification hierarchy-a potential method for integrating image classifiers and expert system rules on geographic context, *International Journal of Remote Sensing*, Vol 11, No 10, 1963-1968.

Woodcock, C. E., and Gopal, S., 2000 Fuzzy set theory and thematic maps: accuracy assessment and area estimation, *International Journal of Geographical Information Science*, Vol 14, 153-172.

Xu, L., Krzyzak, A., Suen, C. Y., 1992 Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition, *IEEE Transactions on Systems, Man and Cybernetics*, Vol 22, 418-435.

Yager, R. R., 1980 An approach to inference in approximate reasoning, *International Journal of Man-Machine Studies*, 13, 323-338.

Yager, R. R., 1984 Approximate Reasoning as a Basis for Rule-Based Expert Systems, *IEEE Transactions on Systems, Man and Cybernetics*, Vol 14, 636-643.

Yoshida, T., and Omatu, S., 1994 Neural Network Approach to Land Cover Mapping, *IEEE Transactions on GeoScience and Remote Sensing*, Vol 32, 1103-1109.

Zadeh, L. A., 1975 Fuzzy Logic and Approximate Reasoning, *Synthese*, 30, 407-428.

Zadeh, L. A., 1965 Fuzzy Sets, *Information and Control*, 8, 338-353.

Zadeh, L. A., 1978 Fuzzy Sets As a Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, 1:3-28.

Zadeh, L. A., 1983 The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems, *Fuzzy Sets and Systems*, 11:199-227.

Zadeh, L. A., 1984 Review of Mathematical theory of evidence, by G Shafer, *AI Magazine*, 5, 81-83.

Zadeh, L. A., 1993 Basic Notions in Fuzzy Set Theory, in *Readings in Fuzzy Sets for Intelligent Systems*, Dubois, D., Prade, H., Yager, R. R., (editors), San Mateo, Calif, Morgan Kaufman Publishers.

Zell, A., Mamier, G., Vogt, M., Mache, N., Hubner, R., Doring, S., Herrmann, K., Soyez, T., Schmalzl, M., Sommer, T., Hatzigeorgiou, A., Posselt, D., Schreiner, T., Kett, B., Clemente, G., Wieland, J., 1995 *Stuttgart Neural Networks Simulator User Manual Version 4.*1, Institute for Parallel and Distributed High Performance System, University of Stuttgart.

Zemankova, M., 1993 FILIP: A Fuzzy Intelligent Information System with Learning Capabilities, in *Readings in Fuzzy Sets for Intelligent Systems*, Dubois, D., Prade, H., Yager, R. R., (editors), San Mateo, Calif, Morgan Kaufman Publishers.

Zeng, T. Q., and Zhou, Q., 2001 Optimal spatial decision making using GIS: a prototype of a real estate geographical information system (REGIS), *International Journal of Geographical Information Science*, Vol 15, 307-321.

Zhang, J. X., and Stuart, N., 2001 Fuzzy methods for categorical mapping with image-based land cover data, *International Journal of Geographical Information Science*, Vol 15, 175-195.

Zhu, A. X., Band, L. E., Dutton, B., Nimlos, T. J., 1996 Automated soil inference under fuzzy logic, *Ecological Modelling*, 90:123-145.

Zhuang, X., Engel, B. A., Baumgardner, M. F., Swain, P. H., 1991 Improving Classification of Crop Residues Using Digital Land Ownership Data and Landsat TM Imagery, *Photogrammetric Engineering & Remote Sensing*, Vol 57, 1487-1492.

Zhuang, X., Engel, B. A., Lozano-Garcia, D. F., Fernandez, R. N., Johannsen, C. J., 1994 Optimization of training data required for neuro-classification, *International Journal of Remote Sensing*, Vol 15, No 16, 3271-3277.

APPENDIX 1

A typical outputs of the user interface of the Fuzzy Expert Systems

```
FuzzyCLIPS> (load "D:/fzclp610b/pc-prjct/borland/type9.txt")
FuzzyCLIPS> (reset)
FuzzyCLIPS> (run 1)
 Welcome to <Kioloa Fuzzy Classification Expert System> version 1.0!

 The copyright is attributed to Mr Zhi Huang, School of Resources,
Environment & Society, Australian National University, 2002!

 Acknowledgements: The system is written using FuzzyCLIPS Version
6.04A (1998) freely provided by NASA!

 Please make sure you have typed in correct words required by the
system (type in '(run 1)' if not otherwise specified!)! Good luck!
FuzzyCLIPS> (run 1)
 Should we start now? (y or n):
y
FuzzyCLIPS> (run 1)
 Choose data input methods! Press '1' for keyboard input, press '2'
for file input.
1
FuzzyCLIPS> (run 1)
 Enter band2 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band2 crisp value (14-101)
 Or Enter 'NP' if you donot have data for band2 value
 or Enter 'why' if you want to know why does the system require this
information:
21
 Enter band4 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band4 crisp value (5-89)
 Or Enter 'NP' if you donot have data for band4 value
 or Enter 'why' if you want to know why does the system require this
information:
why
Landsat TM band4 is often used to discriminate vegetation from water
and soil!

 Enter band4 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band4 crisp value (5-89)
 Or Enter 'NP' if you donot have data for band4 value
 or Enter 'why' if you want to know why does the system require this
information:
29
 Enter band7 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band7 crisp value (0-86)
 Or Enter 'NP' if you donot have data for band7 value
 or Enter 'why' if you want to know why does the system require this
information:
low
 Enter confidence factor for fuzzy band7 value (0.0-1.0):
0.8
 Enter dtm fuzzy value (level, very-low,low,somewhat-low,sligtly-
low,slightly-high,somewhat-high,high,very-high)
 Or Enter dtm crisp value (0-280)
```

```
 Or Enter 'NP' if you donot have data for dtm value
 or Enter 'why' if you want to know why does the system require this
information:
46
 Enter slope fuzzy value (flat, very-gentle,gentle,somewhat-
gentle,sligtly-gentle,slightly-deep,somewhat-deep,deep,very-deep)
 Or Enter slope crisp value (0-36)
 Or Enter 'NP' if you donot have data for slope value
 or Enter 'why' if you want to know why does the system require this
information:
4
 Enter slope aspect value (all-direction,north,north-east,east,east-
south,south,west-south,west,west-noth)
 Or Enter aspect crisp value (-1-360)
 Or Enter 'NP' if you donot have data for aspect value
 or Enter 'why' if you want to know why does the system require this
information:
351
 Enter geology value (sea,quatenary-alluvium,tertiary-
essexite,snapper-point-permian,pebbly-beach-permian,wasp-head-
permian,ordovician, or NP)
 or Enter 'why' if you want to know why does the system require this
information:
pebbly-beach-permian
FuzzyCLIPS> (run 1)
 Type in '(load-facts "fuzzyfact.dat")' to load your data into the
Fuzzy Expert System!

 Then type in '(agenda)' to see how many rules be partially matched by
the facts your import.
FuzzyCLIPS> (load-facts "fuzzyfact.dat")
TRUE
FuzzyCLIPS> (agenda)
1      ask-action: f-7
0      class1rule17: f-8,f-9,f-10,f-11,f-12,f-13,f-14
0      class1rule2: f-8,f-9,f-10,f-11,f-12,f-13,f-14
0      class1rule2: f-8,f-9,f-10,f-11,f-12,f-13,f-14
0      class1rule7: f-8,f-9,f-10,f-11,f-12,f-13,f-14
0      class1rule7: f-8,f-9,f-10,f-11,f-12,f-13,f-14
0      class3rule3: f-8,f-9,f-10,f-11,f-12,f-13,f-14
-10    class10: f-11,f-14,f-12,f-8,f-9,f-10,f-13
For a total of 8 activations.
FuzzyCLIPS> (run 1)
 Please type in '(run x)' which 'x' is the numbers of the
classification rules partially match the facts (except class10). If
only class10 matches the facts, please type in '(run 1)'!

 Do find out which rule give the highest CF!

 Next, please type in '(unwatch all)'!

 Then, please type in '(focus fuzzify)'!

 Finally, please type in '(run)'!
==> f-15    (action display-rule) CF 1.00
<== f-7     (action ask-action) CF 1.00
FuzzyCLIPS> (run 6)
FIRE    1 class1rule17: f-8,f-9,f-10,f-11,f-12,f-13,f-14
==> f-16     (class (value 1) (location 1) (name Dry Sclerophyll)) CF
0.02
FIRE    2 class1rule2: f-8,f-9,f-10,f-11,f-12,f-13,f-14
~CF f-16     (class (value 1) (location 1) (name Dry Sclerophyll)) CF
0.24
FIRE    3 class1rule2: f-8,f-9,f-10,f-11,f-12,f-13,f-14
```

```
~CF f-16     (class (value 1) (location 1) (name Dry Sclerophyll)) CF
0.60
FIRE    4 class1rule7: f-8,f-9,f-10,f-11,f-12,f-13,f-14
FIRE    5 class1rule7: f-8,f-9,f-10,f-11,f-12,f-13,f-14
FIRE    6 class3rule3: f-8,f-9,f-10,f-11,f-12,f-13,f-14
==> f-17      (class (value 3) (location 1) (name Lower slope wet
forest)) CF 0.16
FuzzyCLIPS> (unwatch all)
FuzzyCLIPS> (focus fuzzify)
TRUE
FuzzyCLIPS> (run)
 Which rule gives the highest CF (please key in the rule name):
class1rule2
 Do you want to see the rule (y or n:
y
(defrule MAIN::class1rule2
   (declare (CF 0.99))
   (band2 (band2 very-low) (location ?i))
   (band4 (band4 somewhat-low) (location ?i))
   (band7 (band7 low) (location ?i))
   (elevation (dtm low) (location ?i))
   (gradient (slope very-gentle) (location ?i))
   (or  (aspect (aspect east) (location ?i))
        (aspect (aspect west-north) (location ?i))
        (aspect (aspect west) (location ?i))
        (aspect (aspect west-south) (location ?i))
        (aspect (aspect north) (location ?i))
        (aspect (aspect south) (location ?i))
        (aspect (aspect east-south) (location ?i)))
   (geology  (geology  snapper-point-permian|pebbly-beach-permian|wasp-
head-permian|ordovician) (location ?i))
   =>
   (assert (class (value 1) (name Dry Sclerophyll) (location ?i))))
FuzzyCLIPS> (run 1)
 Type in '(facts)' then press '<cr>' to get the final answer
FuzzyCLIPS> (facts)
f-0      (initial-fact) CF 1.00
f-1      (location 1) CF 1.00
f-3      (action NULL) CF 1.00
f-16     (class (value 1) (location 1) (name Dry Sclerophyll)) CF 0.60
f-18     (geol pebbly-beach-permian) CF 1.00
f-19     (slopeaspect north) CF 1.00
f-22     (slopegradient very-gentle) CF 0.68
f-23     (dem low) CF 1.00
f-25     (7band low) CF 0.80
f-27     (4band somewhat-low) CF 1.00
f-28     (2band very-low) CF 0.61
f-31     (action judge-answer) CF 1.00
For a total of 12 facts.
FuzzyCLIPS> (run 1)
 Is the answer correct (y or n):
y
FuzzyCLIPS> (run 1)
 Thank you! Do you want to continue? (y or n):
y
FuzzyCLIPS> (run 1)
 Choose data input methods! Press '1' for keyboard input, press '2'
for file input.
1
FuzzyCLIPS> (run 1)
 Enter  band2  fuzzy  value  (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band2 crisp value (14-101)
 Or Enter 'NP' if you donot have data for band2 value
```

```
 or Enter 'why' if you want to know why does the system require this
information:
```
20
```
 Enter band4 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band4 crisp value (5-89)
 Or Enter 'NP' if you donot have data for band4 value
 or Enter 'why' if you want to know why does the system require this
information:
```
32
```
 Enter band7 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band7 crisp value (0-86)
 Or Enter 'NP' if you donot have data for band7 value
 or Enter 'why' if you want to know why does the system require this
information:
```
10
```
 Enter dtm fuzzy value (level, very-low,low,somewhat-low,sligtly-
low,slightly-high,somewhat-high,high,very-high)
 Or Enter dtm crisp value (0-280)
 Or Enter 'NP' if you donot have data for dtm value
 or Enter 'why' if you want to know why does the system require this
information:
```
NP
```
 Enter slope fuzzy value (flat, very-gentle,gentle,somewhat-
gentle,sligtly-gentle,slightly-deep,somewhat-deep,deep,very-deep)
 Or Enter slope crisp value (0-36)
 Or Enter 'NP' if you donot have data for slope value
 or Enter 'why' if you want to know why does the system require this
information:
```
4
```
 Enter slope aspect value (all-direction,north,north-east,east,east-
south,south,west-south,west,west-noth)
 Or Enter aspect crisp value (-1-360)
 Or Enter 'NP' if you donot have data for aspect value
 or Enter 'why' if you want to know why does the system require this
information:
```
16
```
 Enter geology value (sea,quatenary-alluvium,tertiary-
essexite,snapper-point-permian,pebbly-beach-permian,wasp-head-
permian,ordovician, or NP)
 or Enter 'why' if you want to know why does the system require this
information:
```
wasp-head-permian

FuzzyCLIPS> *(run 1)*
```
 Type in '(load-facts "fuzzyfact.dat")' to load your data into the
Fuzzy Expert System!

 Then type in '(agenda)' to see how many rules be partially matched by
the facts your import.
```
FuzzyCLIPS> *(load-facts "fuzzyfact.dat")*

```
TRUE
```
FuzzyCLIPS> *(agenda)*
```
1       ask-action: f-36
0       class1rule200: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class1rule28: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class1rule46: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class1rule2: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class1rule7: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class1rule7: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class6rule10: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class6rule3: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class6rule16: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0       class1rule120: f-37,f-38,f-39,f-48,f-49,f-50,f-51
```

```
0      class8rule30: f-37,f-38,f-39,f-48,f-49,f-50,f-51
0      class8rule30: f-37,f-38,f-39,f-48,f-49,f-50,f-51
-10    class10: f-48,f-51,f-49,f-37,f-38,f-39,f-50
For a total of 14 activations.
FuzzyCLIPS> (run 1)
 Please type in '(run x)' which 'x' is the numbers of the
classification rules partially match the facts (except class10). If
only class10 matches the facts, please type in '(run 1)'!

 Do find out which rule give the highest CF!

 Next, please type in '(unwatch all)'!

 Then, please type in '(focus fuzzify)'!

 Finally, please type in '(run)'!
==> f-52    (action display-rule) CF 0.60
<== f-36    (action ask-action) CF 0.60
FuzzyCLIPS> (run 12)
FIRE    1 class1rule200: f-37,f-38,f-39,f-48,f-49,f-50,f-51
==> f-53     (class (value 1) (location 1) (name Dry Sclerophyll)) CF
0.50
FIRE    2 class1rule28: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FIRE    3 class1rule46: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FIRE    4 class1rule2: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FIRE    5 class1rule7: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FIRE    6 class1rule7: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FIRE    7 class6rule10: f-37,f-38,f-39,f-48,f-49,f-50,f-51
==> f-54     (class (value 6) (location 1) (name Rainforest Ecotone))
CF 0.01
FIRE    8 class6rule3: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FIRE    9 class6rule16: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FIRE   10 class1rule120: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FIRE   11 class8rule30: f-37,f-38,f-39,f-48,f-49,f-50,f-51
==> f-55     (class (value 8) (location 1) (name clear land or
paddock)) CF 0.01
FIRE   12 class8rule30: f-37,f-38,f-39,f-48,f-49,f-50,f-51
FuzzyCLIPS> (unwatch all)
FuzzyCLIPS> (focus fuzzify)
TRUE
FuzzyCLIPS> (run)
 Which rule gives the highest CF (please key in the rule name):
class1rule200
 Do you want to see the rule (y or n):
y
(defrule MAIN::class1rule200
   (declare (CF 0.99))
   (band2 (band2 very-low) (location ?i))
   (band4 (band4 somewhat-low) (location ?i))
   (band7 (band7 low) (location ?i))
   (elevation (dtm low) (location ?i))
   (gradient (slope very-gentle) (location ?i))
   (aspect (aspect north-east) (location ?i))
   (geology (geology pebbly-beach-permian|wasp-head-permian) (location
?i))
   =>
   (assert (class (value 1) (name Dry Sclerophyll) (location ?i))))
FuzzyCLIPS> (run 1)
 Type in '(facts)' then press '<cr>' to get the final answer
FuzzyCLIPS> (facts)
f-0      (initial-fact) CF 1.00
f-1      (location 1) CF 1.00
f-3      (action NULL) CF 1.00
f-53     (class (value 1) (location 1) (name Dry Sclerophyll)) CF 0.50
```

```
f-56     (geol wasp-head-permian) CF 1.00
f-58     (slopeaspect north) CF 0.97
f-60     (slopegradient very-gentle) CF 0.68
f-69     (dem level) CF 0.50
f-71     (7band low) CF 1.00
f-72     (4band somewhat-low) CF 1.00
f-73     (2band very-low) CF 0.80
f-76     (action judge-answer) CF 0.60
For a total of 12 facts.
FuzzyCLIPS> (run 1)
 Is the answer correct (y or n):
y
FuzzyCLIPS> (run 1)
 Thank you! Do you want to continue? (y or n):
y
FuzzyCLIPS> (run 1)
 Choose data input methods! Press '1' for keyboard input, press '2'
for file input.
1
FuzzyCLIPS> (run 1)
 Enter    band2    fuzzy    value    (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band2 crisp value (14-101)
 Or Enter 'NP' if you donot have data for band2 value
 or Enter 'why' if you want to know why does the system require this
information:
24
 Enter    band4    fuzzy    value    (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band4 crisp value (5-89)
 Or Enter 'NP' if you donot have data for band4 value
 or Enter 'why' if you want to know why does the system require this
information:
50
 Enter    band7    fuzzy    value    (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band7 crisp value (0-86)
 Or Enter 'NP' if you donot have data for band7 value
 or Enter 'why' if you want to know why does the system require this
information:
14
 Enter   dtm   fuzzy   value   (level,   very-low,low,somewhat-low,sligtly-
low,slightly-high,somewhat-high,high,very-high)
 Or Enter dtm crisp value (0-280)
 Or Enter 'NP' if you donot have data for dtm value
 or Enter 'why' if you want to know why does the system require this
information:
158
 Enter   slope   fuzzy   value   (flat,   very-gentle,gentle,somewhat-
gentle,sligtly-gentle,slightly-deep,somewhat-deep,deep,very-deep)
 Or Enter slope crisp value (0-36)
 Or Enter 'NP' if you donot have data for slope value
 or Enter 'why' if you want to know why does the system require this
information:
9
 Enter  slope  aspect  value  (all-direction,north,north-east,east,east-
south,south,west-south,west,west-noth)
 Or Enter aspect crisp value (-1-360)
 Or Enter 'NP' if you donot have data for aspect value
 or Enter 'why' if you want to know why does the system require this
information:
89
```

```
 Enter    geology    value    (sea,quatenary-alluvium,tertiary-
essexite,snapper-point-permian,pebbly-beach-permian,wasp-head-
permian,ordovician, or NP)
 or Enter 'why' if you want to know why does the system require this
information:
ordovician
FuzzyCLIPS> (run 1)
 Type in '(load-facts "fuzzyfact.dat")' to load your data into the
Fuzzy Expert System!

 Then type in '(agenda)' to see how many rules be partially matched by
the facts your import.
FuzzyCLIPS> (load-facts "fuzzyfact.dat")
TRUE
FuzzyCLIPS> (agenda)
1      ask-action: f-81
-10    class10: f-85,f-88,f-86,f-82,f-83,f-84,f-87
For a total of 2 activations.
FuzzyCLIPS> (run 1)
 Please type in '(run x)' which 'x' is the numbers of the
classification rules partially match the facts (except class10). If
only class10 matches the facts, please type in '(run 1)'!

 Do find out which rule give the highest CF!

 Next, please type in '(unwatch all)'!

 Then, please type in '(focus fuzzify)'!

 Finally, please type in '(run)'!
==> f-89    (action display-rule) CF 0.50
<== f-81    (action ask-action) CF 0.50
FuzzyCLIPS> (run 1)
FIRE    1 class10: f-85,f-88,f-86,f-82,f-83,f-84,f-87
==> f-90    (class (value 10) (location 1) (name unknown)) CF 1.00
FuzzyCLIPS> (unwatch all)
FuzzyCLIPS> (focus fuzzify)
TRUE
FuzzyCLIPS> (run)
 Which rule gives the highest CF (please key in the rule name):
class10
 Do you want to see the rule (y or n):
y
(defrule MAIN::class10
   (declare (salience -10))
   ?elevation <- (elevation (location ?i) (dtm ?a))
   ?geology <- (geology (location ?i) (geology ?c))
   ?gradient <- (gradient (location ?i) (slope ?b))
   ?band2 <- (band2 (location ?i) (band2 ?d))
   ?band4 <- (band4 (location ?i) (band4 ?e))
   ?band7 <- (band7 (location ?i) (band7 ?f))
   ?aspect <- (aspect (location ?i) (aspect ?g))
   =>
   (assert (class (value 10) (name unknown) (location ?i))))
FuzzyCLIPS> (run 1)
 Type in '(facts)' then press '<cr>' to get the final answer
FuzzyCLIPS> (facts)
f-0     (initial-fact) CF 1.00
f-1     (location 1) CF 1.00
f-3     (action NULL) CF 1.00
f-90    (class (value 10) (location 1) (name unknown)) CF 1.00
f-91    (geol ordovician) CF 1.00
f-93    (slopeaspect east) CF 1.00
f-95    (slopegradient somewhat-gentle) CF 0.51
```

```
f-96     (dem slightly-high) CF 1.00
f-97     (7band low) CF 1.00
f-98     (4band slightly-strong) CF 1.00
f-100    (2band low) CF 1.00
f-102    (action judge-answer) CF 0.50
For a total of 12 facts.
FuzzyCLIPS> (run 1)
 Is the answer correct (y or n):
n
FuzzyCLIPS> (run 1)
 What is right answer anyway? (1-10):
5

 Thank you! I have learned the rule from you! Do you want to continue?
(y or n):
y
FuzzyCLIPS> (run 1)
 Choose data input methods! Press '1' for keyboard input, press '2'
for file input.
1
FuzzyCLIPS> (run 1)
 Enter band2 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band2 crisp value (14-101)
 Or Enter 'NP' if you donot have data for band2 value
 or Enter 'why' if you want to know why does the system require this
information:
24
 Enter band4 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band4 crisp value (5-89)
 Or Enter 'NP' if you donot have data for band4 value
 or Enter 'why' if you want to know why does the system require this
information:
50
 Enter band7 fuzzy value (very-low,low,somewhat-low,sligtly-
low,slightly-strong,somewhat-strong,strong,very-strong)
 Or Enter band7 crisp value (0-86)
 Or Enter 'NP' if you donot have data for band7 value
 or Enter 'why' if you want to know why does the system require this
information:
14
 Enter dtm fuzzy value (level, very-low,low,somewhat-low,sligtly-
low,slightly-high,somewhat-high,high,very-high)
 Or Enter dtm crisp value (0-280)
 Or Enter 'NP' if you donot have data for dtm value
 or Enter 'why' if you want to know why does the system require this
information:
158
 Enter slope fuzzy value (flat, very-gentle,gentle,somewhat-
gentle,sligtly-gentle,slightly-deep,somewhat-deep,deep,very-deep)
 Or Enter slope crisp value (0-36)
 Or Enter 'NP' if you donot have data for slope value
 or Enter 'why' if you want to know why does the system require this
information:
9
 Enter slope aspect value (all-direction,north,north-east,east,east-
south,south,west-south,west,west-noth)
 Or Enter aspect crisp value (-1-360)
 Or Enter 'NP' if you donot have data for aspect value
 or Enter 'why' if you want to know why does the system require this
information:
89
```

```
 Enter     geology     value     (sea,quatenary-alluvium,tertiary-
essexite,snapper-point-permian,pebbly-beach-permian,wasp-head-
permian,ordovician, or NP)
 or Enter 'why' if you want to know why does the system require this
information:
ordovician
FuzzyCLIPS> (run 1)
 Type in '(load-facts "fuzzyfact.dat")' to load your data into the
Fuzzy Expert System!

 Then type in '(agenda)' to see how many rules be partially matched by
the facts your import.
FuzzyCLIPS> (load-facts "fuzzyfact.dat")
TRUE
FuzzyCLIPS> (agenda)
1      ask-action: f-107
0       class5newrule518570.715: f-108,f-109,f-110,f-111,f-112,f-113,f-
114
-10    class10: f-111,f-114,f-112,f-108,f-109,f-110,f-113
For a total of 3 activations.
FuzzyCLIPS> (run 1)
 Please  type  in  '(run  x)'  which  'x'  is  the  numbers  of  the
classification rules partially match the facts (except class10). If
only class10 matches the facts, please type in '(run 1)'!

 Do find out which rule give the highest CF!

 Next, please type in '(unwatch all)'!

 Then, please type in '(focus fuzzify)'!

 Finally, please type in '(run)'!
==> f-115   (action display-rule) CF 0.50
<== f-107   (action ask-action) CF 0.50
FuzzyCLIPS> (run 1)
FIRE     1  class5newrule518570.715:  f-108,f-109,f-110,f-111,f-112,f-
113,f-114
==> f-116    (class  (value  5)  (location  1)  (name  dry  E.maculata))  CF
0.51
FuzzyCLIPS> (unwatch all)
FuzzyCLIPS> (focus fuzzify)
TRUE
FuzzyCLIPS> (run)
 Which rule give the highest CF (please key in the rule name):
class5newrule518570.715
 Do you want to see the rule (y or n):
y
(defrule MAIN::class5newrule518570.715
   (band2 (band2 low) (location ?i))
   (band4 (band4 slightly-strong) (location ?i))
   (band7 (band7 low) (location ?i))
   (elevation (dtm slightly-high) (location ?i))
   (gradient (slope somewhat-gentle) (location ?i))
   (aspect (aspect east) (location ?i))
   (geology (geology ordovician) (location ?i))
   =>
   (assert (class (value 5) (name dry E.maculata) (location ?i))))
FuzzyCLIPS> (run 1)
 Type in '(facts)' then press '<cr>' to get the final answer
FuzzyCLIPS> (facts)
f-0      (initial-fact) CF 1.00
f-1      (location 1) CF 1.00
f-3      (action NULL) CF 1.00
f-116    (class (value 5) (location 1) (name dry E.maculata)) CF 0.51
```

```
f-117    (geol ordovician) CF 1.00
f-119    (slopeaspect east) CF 1.00
f-121    (slopegradient somewhat-gentle) CF 0.51
f-122    (dem slightly-high) CF 1.00
f-123    (7band low) CF 1.00
f-124    (4band slightly-strong) CF 1.00
f-126    (2band low) CF 1.00
f-128    (action judge-answer) CF 0.50
For a total of 12 facts.
FuzzyCLIPS> (run 1)
 Is the answer correct (y or n):
y
FuzzyCLIPS> (run 1)
 Thank you! Do you want to continue? (y or n):
n
FuzzyCLIPS> (exit)
```